Environment Agency

Evidence

Infaunal quality index: Water Framework Directive classification scheme for marine benthic invertebrates

Report: SC080016

The Environment Agency is the leading public body protecting and improving the environment in England.

It's our job to make sure that air, land and water are looked after by everyone in today's society, so that tomorrow's generations inherit a cleaner, healthier world.

Our work includes tackling flooding and pollution incidents, reducing industry's impacts on the environment, cleaning up rivers, coastal waters and contaminated land, and improving wildlife habitats.

This report is the result of research commissioned and funded by the Environment Agency.

Published by:
Environment Agency, Horizon House, Deanery Road, Bristol, BS1 5AH
www.environment-agency.gov.uk

ISBN: 978-1-84911-319-9

E: enquiries@environment-agency.gov.uk.

Author(s):
G. R. Phillips, A. Anwar, L. Brooks, L. J. Martina, A. C. Miles, A. Prior

Dissemination Status:
Publicly available

Keywords:
Water Framework Directive, Infaunal Quality Index, benthic invertebrates, ecological classification

Research Contractor:
Graham Phillips, Environment Agency, Kingfisher House, Goldhay Way, Peterborough, PE2 5ZR Tel: 01733 464229

Environment Agency's Project Manager:
Graham Phillips, National Operations Directorate

Collaborator(s):
Northern Ireland Environment Agency
Scottish and Northern Ireland Forum for Environmental Research
Scottish Environment Protection Agency

Project Number:
SC080016

Evidence at the Environment Agency

Evidence underpins the work of the Environment Agency. It provides an up-to-date understanding of the world about us, helps us to develop tools and techniques to monitor and manage our environment as efficiently and effectively as possible. It also helps us to understand how the environment is changing and to identify what the future pressures may be.

The work of the Environment Agency's Evidence Directorate is a key ingredient in the partnership between research, guidance and operations that enables the Environment Agency to protect and restore our environment.

This report was produced by the Scientific and Evidence Services team within Evidence. The team focuses on four main areas of activity:

- **Setting the agenda**, by providing the evidence for decisions;
- **Maintaining scientific credibility**, by ensuring that our programmes and projects are fit for purpose and executed according to international standards;
- **Carrying out research**, either by contracting it out to research organisations and consultancies or by doing it ourselves;
- **Delivering information, advice, tools and techniques**, by making appropriate products available.

Miranda Kavanagh

Director of Evidence

Executive summary

The Infaunal Quality Index (IQI) was developed to assess the ecological status of the macrobenthic invertebrate infaunal assemblages of sediment habitats in UK coastal and transitional water bodies for the Water Framework Directive (WFD). This report follows on from the initial phases documented in the interim technical report published in 2004 and describes the development of the IQI as used for assessment in the first UK River Basin Management Plans (RBMPs). This report document the development of version IV of the IQI formula.

The IQI is a multimetric index that expresses the ecological health of benthic macroinvertebrate (infauna) assemblages in accordance with the normative definitions (Annex V, the basis upon which ecological status is defined) of the WFD as an Ecological Quality Ratio (EQR). The multimetric was developed to reflect how the structure and functioning of benthic macroinvertebrate assemblage change over anthropogenic pressure gradients. A subset of WFD compliant metrics were selected that would, when used in combination, encompass a high amount of information on how macroinvertebrate assemblages change within the marine environment. The selected metrics were taxa number, the AZTI Marine Biotic Index (AMBI, a measure of sensitivity to disturbance) and Simpson's evenness (a measure of the distribution of individuals across the different taxa). To fulfil the requirements of the WFD, the $IQI_{v.IV}$ incorporates each metric as a ratio of the observed value to that expected under reference conditions.

The correlation between the metric subset and contaminant data from two pressure gradient datasets (organic enrichment and hazardous substances/physical smothering) was used to derive the extent to which each metric should contribute to $IQI_{v.IV}$.

Values of taxa number, AMBI and Simpson's evenness expected under reference conditions were established. Values were originally set according to expert judgement from representatives of the competent authorities of the UK and Republic of Ireland for a limited range of marine sediments and sampling methods. The correlation between the $IQI_{v.IV}$ metrics and different environmental factors was used to expand the range of environmental conditions to which reference conditions (and therefore $IQI_{v.IV}$) could be applied.

Ecological status boundaries were provisionally set at the national level to ensure the proportion of individuals in different sensitivity groups (AMBI ecological groups) corresponded to those expected according to the WFD normative definitions. National boundaries were subsequently adjusted to maximise agreement between status classes as derived by the classification methods of the North East Atlantic Geographical Intercalibration Group Member States. These boundaries were used in the first RBMPs.

Development of the IQI and associated reference conditions depended on the quality and range of the data available. The data treatment rules established in the interim technical report (2004) to standardise soft sediment benthic macroinvertebrate datasets have been revised to account for improvements in taxonomic identification and to allow the inclusion of taxa from a broader range of habitats in the assessments.

An approach to estimating EQR variability for the $IQI_{v.IV}$ was derived in collaboration between the Environment Agency and the Water Research Centre (WRc). The variability estimates provide the basis for establishing statistical confidence in ecological status of an assessment (confidence of class or risk of misclassification) and in establishing the appropriate sample effort for the WFD monitoring programme.

Acknowledgements

The authors wish to acknowledge the following for their contributions and feedback into the work underlying the present report: members of the Marine Task Team and Marine Benthic Invertebrate Task Team of the United Kingdom and Republic of Ireland, in particular Tim Mackie (Northern Ireland Environment Agency), Myles O' Reilly (Scottish Environment Protection Agency), Carol Milner (Scottish Environment Protection Agency), Francis O'Beirn (Marine Institute of the Republic of Ireland), Yvonne Leahy (Marine Institute of the Republic of Ireland), Lucie Skates (Countryside Council for Wales) and Oliver Avis-Crawford (Environment Agency).

The authors also wish to acknowledge AstraZeneca UK Limited and Fisheries Research Services for contributing data key to the work in the underlying report.

Bob Clarke, Richard Warwick and Paul Somerfield (Plymouth Marine Laboratories) must be acknowledged for their contribution to the development of the Infaunal Quality Index and Julian Ellis (Water Research Centre) for his invaluable contribution to the approach for assessing the statistical certainty of the assessment method.

Acknowledgments must also go to the reviewers Paul Somerfield (Plymouth Marine Laboratories), Michael Elliott (Institute of Estuarine and Coastal Studies, University of Hull) and Silvana Birchenough (Cefas) for their valuable feedback prior to the completion of this report.

The project was part funded by the Scottish and Northern Ireland Forum for Environmental Research.

Contents

1 Introduction

1.1 Background to the Water Framework Directive

The Water Framework Directive (2000/60/EC) substantially altered the approach to water management in Europe, establishing a framework for the protection of all waters (inland surface waters, transitional waters, coastal waters and groundwater). It aims to integrate environmental assessments (disciplines, analyses and expertise) within river basin catchment areas to give 'joined up' management of the water environment (European Parliament and Council of the European Union 2000).

With regard to the marine environment, the main purposes of the Water Framework Directive (WFD) are to:

- prevent deterioration and protect and enhance the status of aquatic ecosystems and associated wetlands
- promote sustainable water use
- reduce pollution from priority substances
- protect territorial and marine waters

Further information on the WFD can be found on the European Commission's WFD web pages (http://ec.europa.eu/environment/water/water-framework/).

The UK Technical Advisory Group for the Water Framework Directive (UKTAG) (http://www.wfduk.org) was established as a forum for providing technical support to the UK administrations and the designated competent monitoring authorities during the transposition and implementation of the WFD.

The intention of the WFD is to restore all inland, transitional and coastal waters to good status by 2015 (Article 4(a)(ii), 2000/60/EC), ensuring that there is no deterioration of ecological status. To determine the overall status of defined areas, the WFD incorporates an ecological status assessment (biological, hydromorphological and physicochemical) and a chemical status assessment. The status assessment is carried out on 'water bodies'. Under Article 5 of the WFD, prior to status assessment, waters were required to be divided in to units or 'water bodies' and characterised according to their type specific conditions and the significant pressures acting on them. Guidance on the implementation of WFD Article 5 was produced on characterisation and typology (UKTAG 2003a) and pressures and impacts analysis (UKTAG 2003b).

To help standardise the WFD a large number of groups at UK and European level were set up to, among other tasks, define and assess 'Good Ecological Status' (GES) and 'Good Chemical Status'. Water bodies designated as heavily modified (HMWB) as a result of hydromorphological pressures are assessed in terms of ecological potential, where the objective of GES is replaced by 'Good Ecological Potential' (GEP). Methods for assessing GEP combine GES classification systems with additional methods.

At the European scale, a common understanding of the WFD requirements such as on the establishment of water body types (typology), reference conditions and classification systems was produced by the Common Implementation Strategy (CIS) COAST Working Group 2.4 (Vincent et al. 2002, COAST 2003). The WFD competent authorities in the UK pooled resources with the Republic of Ireland (RoI), forming the

UK and RoI Marine Task Team (MTT) to coordinate the adaptation and development of suitable classification tools for the biological quality elements for coastal and transitional waters, including the production of guidance. Overseen by UKTAG, MTT set up expert groups (Figure 1.1) to develop WFD compliant classification tools for all the biological quality elements, that is:

- angiosperms
- benthic invertebrates
- macroalgae
- phytoplankton
- transitional water fish

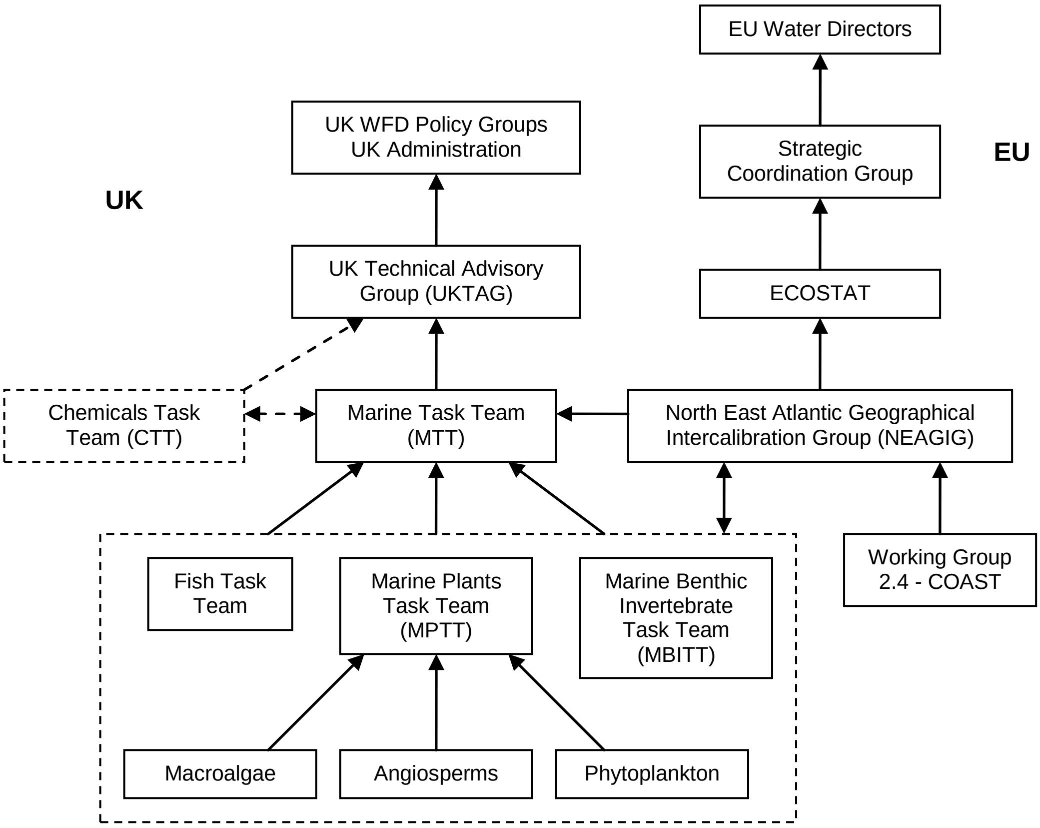


Figure 1.1 Groups involved in the development of UK ecological classification tools for transitional and coastal waters

For each biological quality element, classification tools are necessary in order to give a statistically robust definition of the 'health' of the element in the defined water body. Under the WFD, 'health' is assessed by comparing the measured conditions against those described for reference (minimally impacted) conditions. The ecological status assessment of a water body therefore relies on defining appropriate reference conditions, which describe the optimum ecological status for a defined water body type, and having a classification assessment method that can quantify deviation from that reference condition. In addition, in order to contribute to effective management of the water environment, classification tools have to show a clear, measurable response to anthropogenic pressure. The measurement of chemical and ecological status through suitable classification tools determines whether the requirements of the WFD are being met and drives a programme of measures (WFD Article 11) should status be identified as ‘less than good’.

1.2 Aim of the report

The aim of this report is to document the development of the benthic invertebrate classification tool, the Infaunal Quality Index (IQI). The IQI was developed to fulfil the requirements of the WFD with respect to the benthic invertebrate quality element in transitional and coastal (TraC) waters.

1.3 Project background

The Marine Benthic Invertebrate Task Team (MBITT) was set up by MTT to develop an integrated approach to assessing the benthic invertebrate component across the UK and the Republic of Ireland. The MBITT project board consisted of technical specialists from the following UK and RoI WFD competent authorities, conservation agencies and research institutions:

- Environment Agency (lead agency)
- Scottish Environment Protection Agency (SEPA)
- Northern Ireland Environment Agency (NIEA)
- Marine Institute, Republic of Ireland (MI RoI)
- Joint Nature Conservation Committee (JNCC)
- Natural England (formerly English Nature)
- Countryside Council for Wales (CCW)
- Centre for Environment Fisheries and Aquaculture Science (Cefas)
- Institute of Estuarine and Coastal Studies (IECS)

In addition, links for external consultation and review were established with marine benthic ecologists from academic, government and consultant organisations.

The work of the project developed alongside guidance from MTT, UKTAG, European Commission and experts from the Member States of the North East Atlantic Geographical Intercalibration Group (NEAGIG)[1], as well as with specific direction from the UK’s WFD competent authorities. As such, the project has contributed to a wide range of tasks such as typology, pressure assessment, data collation and quality issues, intercalibration, monitoring strategies and sampling procedures, some of which have had to be revisited as understanding of the WFD has developed.

The development of the IQI involved four stages (see Chapter 3). The early phases (2001-2004) of the development of a WFD classification scheme for the marine benthic invertebrate component are documented in the interim technical report (Prior et al. 2004), which describes work under Environment Agency projects E1-116 and E1-132.

[1] Belgium, Denmark, France, Germany, Ireland, the Netherlands, Norway, Portugal, Spain, Sweden and the UK.

Following the initial stages, development focused on the work areas initiated under phase III of Environment Agency project E1-139 (Prior et al. 2004):

- refinement of the IQI to ensure response to anthropogenic pressure
- continued collation, quality assurance and storage of WFD compliant macrobenthic invertebrate abundance data
- establishment of habitat-specific reference conditions
- quantification of the variability of ecological status assessment, as shown through classification tools, based on the natural variability of benthic communities
- quantification of the risk of misclassification of ecological assessment using the selected tools
- comparison with benthic invertebrate classification tools being developed in other North East Atlantic Member States (NEAGIG intercalibration process).
- development of water body assessment protocols

Staged updates on the project have been provided through internal reports and contributions to the North East Atlantic Geographical Intercalibration Group (for example, Borja et al. 2007, Carletti and Heiskanen 2009).

This report documents the technical development of the IQI (version IV) used for the assessment of the benthic invertebrate component of UK transition and coastal waters reported in the first River Basin Management Plans (RBMPs)[2] in 2009. The legally defined use of the IQI within these plans for the UK is specified by:

- The River Basin Districts Typology, Standards and Groundwater Threshold Values (Water Framework Directive) (England and Wales) Directions 2009
- The Scotland River Basin District (Surface Water Typology, Environmental Standards, Condition Limits and Groundwater Threshold Values) Directions 2009
- The Water Framework Directive (Priority Substances and Classification) Regulations (Northern Ireland) 2011
- UKTAG coastal water assessment method statement for benthic invertebrate fauna (UKTAG 2008)

1.4 Structure of the report

The work carried out during the project can be divided into four main areas:

[2]RBMPs are plans for protecting and improving the water environment. They present the main issues for the water environment and the actions needed to deal with them. For more information about how the plans are produced see the Environment Agency's web page on River Basin Management Planning (http://www.environment-agency.gov.uk/research/planning/33240.aspx). The RBMPs themselves can be viewed on the Environment Agency's web page on managing and improving the water environment (http://www.environment-agency.gov.uk/research/planning/33106.aspx).

- the IQI formula
- reference conditions
- status class boundaries
- data treatment

Figure 1.2 shows a timeline of development for these four work areas. To provide clear documentation of the IQI, the work is further delineated and presented in the report as follows:

- **Data treatment protocols (Chapter 2) –** the development of protocols to standardise the data used for tool development, reference conditions, intercalibration and classification
- **Classification tool development (Chapter 3)** – the approach taken to identify, adapt and combine indicators of biological health into a single index responsive to anthropogenic pressures
- **Setting reference conditions (Chapter 4)** – the approach taken to set metric values expected under minimally impacted conditions, and their adaptation in response to changing natural environmental conditions
- **Setting ecological status class boundaries (Chapter 5)** – the approach taken in interpreting the WFD normative definitions into ecological status class boundaries for the UK and RoI, and their adaptation to align with classification systems across the Member States of the North East Atlantic Geographical Intercalibration Group
- **Estimating Ecological Quality Ratio variability (Chapter 6)** – the approach taken to estimate Ecological Quality Ratio (EQR) variability
- **Estimating the risk of misclassification (Chapter 7)** – the approach whereby data variability is used to estimate the statistical likelihood that an incorrect ecological status has been assigned to a classification
- **Power analysis (Chapter 8)** – the approach taken to use EQR variability to estimate the sampling effort necessary to support the design of WFD monitoring programmes

	2004	2005	2006	2007	2008	2009
IQI formula:	IQI v.I developed IQI v.II developed	IQI v.III developed	IQI v.IV developed			
Reference conditions:		Reference conditions incorporated within IQI v.III	Agreed for marine sublittoral fine sand/mud ($0.1m^2$, 1mm)		Reference conditions revised due to truncation rule changes Reference condition continuous habitat approach developed	
Status class boundaries:		UK status class boundaries set	Boundaries revised through intercalibration (phase I) for coastal waters	Revision of optimised boundaries as a result of updated UK reference conditions		
Data treatment rules:	Data treatment rules for fine sand/mud habitats established in phase II				Data treatment rules adapted for a broader range of habitats and analysis improvements	

Figure 1.2 Timeline for the development of the IQI classification tool with associated reference conditions, status class boundaries and data treatment rules

While the project work areas are defined here as discrete topics, there is a high level of interaction between the different areas (for example, adapting the reference conditions has implications for the status class boundaries). Over the course of defining the IQI, each work area has undergone multiple revisions as a result of the development of methods, recommendations through consultation and the availability of additional data. The interdependency between the different work areas has meant that developments in one area often had consequences for other areas (Figure 1.3). For example, modifications in the approach to data treatment influence the setting of reference conditions and the behaviour of the ecological quality ratio (EQR) values, in turn affecting the placement of the status class boundaries.

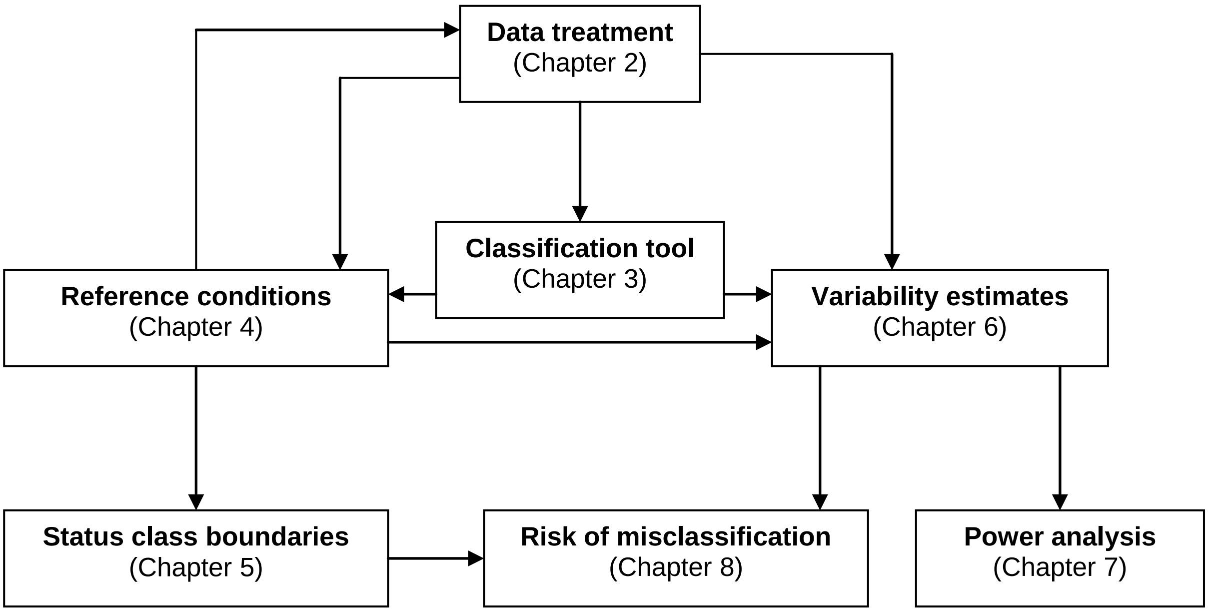


Figure 1.3 Interactions between the different work areas of the overall process of developing the classification tool

1.5 Future developments

This report describes the development of the IQI for use in the first UK River Basin Management Plans. However, further revision may occur through the continuing WFD process. As WFD-specific data are collected through national monitoring programmes, further pressures will be identified and class status boundaries further intercalibrated (NEAGIG Intercalibration Phase 2). The performance of the classification tools will therefore be reviewed to ensure they give us effective tools for managing anthropogenic pressures in the UK's transitional and coastal waters.

2 Data treatment protocols

The development of the IQI with associated reference conditions is heavily dependent upon the quality and range of the data available. The process has relied on a broad range of spatial and temporal data from a multitude of sources. Biological data are inherently variable, but variability also arises through sample collection, processing and analysis. This chapter describes the approach taken to standardising benthic invertebrate data for the purposes of tool development, the setting of reference conditions and classification.

2.1 Background

Before any data analysis and classification tool development, it is vital to ensure data standardisation. If data are not carefully managed, apparent variations in ecological class status may arise as an artefact of data recording. Throughout the project the treatment of data, whether from UK agencies or the North East Atlantic Geographical Intercalibration Group, has been pivotal for the success of the other work areas (see Section 1.4).

The WFD acknowledges the potential for discrepancies within ecological data. Annex V, section 1.3.6, requires that:

> 'Methods used for the monitoring of type parameters shall conform to the international standards listed [in section 1.3.6] or such other national or international standards which will ensure the provision of data of an equivalent scientific quality and comparability'.

The UK WFD competent monitoring authorities follow the ISO/CEN standards for benthic invertebrate sampling and processing (ISO 2005).

Ensuring data comparability requires controls to be in place to mitigate, where possible, the effect of any factor that may be a potential source of discrepancy between data. It is important that data comparability is considered at the monitoring design stage where spatial (for example, proximity to pressures, salinity or substratum considerations), temporal (for example, seasonality) and methodological (for example, sieve mesh, sample collection method) aspects may reduce the comparability if appropriate controls are not in place.

Quality assurance at the survey planning and sample collection stage may be upheld using monitoring protocols appropriate to the purpose of the survey such as the *Marine Monitoring Handbook* (Davies et al. 2001), the *Green Book* produced by the Clean Seas Environmental Monitoring Programme (CSEMP) (Cefas 2009) and WFD Operational Instructions drawn up by the Environment Agency, SEPA and NIEA.

Following sample collection, the identification and enumeration of taxa within macrofauna datasets can still be subject to a range of inconsistencies from a variety of sources. These inconsistencies have the potential to reduce the quality and comparability of data used for the purpose of tool development, classification and setting reference conditions. Inconsistencies may result from:

- variation in use of identification literature
- varying degrees of taxonomic specialisation of analysts

- differing identification protocols (for example, the taxonomic level to which certain taxa are identified, classification of juveniles and so on)
- differing enumeration protocols (for example, subsampling methods, treatment of colonial taxa)
- variation in (or absence of) quality assurance protocols
- varying approaches to the exclusion of certain taxonomic groups (for example, epifaunal taxa, pelagic taxa)

These discrepancies, which may occur in data at stages from laboratory analysis to data storage and reporting, all have the potential to alter metric values derived from the data, resulting in under or overestimation of such values within the classification tool.

Within the UK, the National Marine Biological Analytical Quality Control (NMBAQC) scheme (www.nmbaqcs.org) was established in 1992-1993 to develop nationally standardised quality control procedures for macrobenthic analysis. Part of the UK commitment to the Oslo and Paris (OSPAR) convention Joint Assessment Monitoring Programme (JAMP), the NMBAQC scheme has been continually developed, resulting in improved accuracy and standardisation in benthic identification and enumeration. However, even with data delivered under analytical quality control (AQC) schemes, it has been necessary to develop data treatment protocols to ensure full standardisation of data when combining present day and historic data from a variety of sources.

2.2 Macrobenthic invertebrate data standardisation (2004)

An approach to data standardisation was initially developed following consultation with experts at a WFD benthic invertebrate workshop held in October 2003 (see Prior et al. 2004). The rules were established in order to address inconsistencies in taxa identification in UK historic datasets used for classification tool development. At that time, classification development focused on a limited range of habitats described by two European Nature Information System (EUNIS) classes (Davies et al. 2004):

- A5.2 marine sublittoral sand and muddy sand
- A5.3 marine sublittoral mud[3]

Data treatment rules for the raw data were developed specifically to standardise data from these habitats for use with the Infaunal Quality Index. These rules can be summarised as follows:

- removal of epifauna
- removal of juveniles
- removal of meiofauna
- removal of non-invertebrate taxa
- removal of freshwater taxa (including insects)

[3] The EUNIS classification system was updated in 2004. EUNIS A5.2 and A5.3 habitats correspond to A4.2 and A4.3 respectively in Prior et al. (2004).

- removal of taxa identified at Phylum level (some exclusions)
- combination of taxa from groups with identification inconsistencies

The 2004 data treatment rules along with their justification are detailed in the interim technical report (Prior et al. 2004).

2.3 Revised macrobenthic invertebrate standardisation rules (2008)

Since the initial stages of the WFD assessment, developments in laboratory analysis (improved application of standardised analytical quality control protocols to data) and changes in how the data are used within the IQI have taken place. In addition, with developments in how reference conditions are set (see Chapter 4), the IQI has been adapted to apply to a broader range of habitats than those classified as EUNIS A5.2 and A5.3 to include intertidal habitats, those exposed to low/variable salinity and those consisting of coarser and mixed sediments. This has had implications for the original data standardisation protocols. In accordance with these changes, taxon standardisation rules were revised in 2008 (Table 2.1) following discussions with MBITT and NMBAQC.

The application of standardisation rules to assemblage data has implications for the associated indices, and consequentially, reference conditions. In addition to the direct influence on taxon number as a result of ungrouping, Simpson's evenness is influenced by the separation of taxa previously grouped at higher levels (for example, Oligochaeta). Grouping of taxa (in relation to the original standardisation rules) distributes the overall abundance over fewer numbers of taxa, with the effect of generally reducing the apparent evenness of the sample. To ensure that IQI values, and associated ecological status boundaries remain relatively constant regardless of adaptations to standardisation rules, reference conditions need to be revised accordingly.

Table 2.1 Summary of standardisation rules developed in 2004 (Prior et al. 2004), with amendments adopted in 2008

Protocol	Justification	2004 Decision	2008 Decision	Comments
Remove juveniles	Occasional high irregularity in spatial and temporal distribution Difficulties with identification	✓	✗	Juveniles of certain taxa may be significant indicators of the ecological health of an ecosystem Preliminary data analysis indicates no significant difference [1] between IQI scores for samples where juveniles are included/excluded (reference values adjusted accordingly)
Remove epifauna	Taxa not representative of EUNIS A5.2 and A5.3 assemblages	✓	✗	Rule modified. IQI adapted for mixed/coarse soft substratum assemblages. Epifaunal taxa may be representative indicators of ecological health

Protocol	Justification	2004 Decision	2008 Decision	Comments
	Inconsistencies in enumeration of colonial taxa			for mixed/coarse substrata. Colonial taxa to be enumerated as 1 (acknowledge presence)
Combine taxa with identification inconsistencies	Insufficient identification and QA protocols	✓	✗	Broad application of standardised AQC protocols for WFD sample analysis with improving consistency in identification of previously problematic taxa
Remove freshwater taxa	Taxa unrepresentative of transitional and coastal assemblages	✓	✗	Uncertainty over whether certain taxa are exclusively freshwater Instances where freshwater taxa are found in TraC samples in high abundance
Remove planktonic taxa	Taxa unrepresentative of benthic assemblages	✓	✓	–
Remove non-invertebrate taxa	Taxa unrepresentative of invertebrate quality element	✓	✓	–
Remove class Insecta	Taxa unrepresentative of TraC benthic invertebrate assemblages	✓	✗	Certain Insecta taxa representative of TraC benthic invertebrate assemblages
Remove meiofauna	Potential bias to assemblages based on high abundance of meiofauna	✓	✗	Infrequent instances of cases where assemblages biased by high meiofaunal abundance

Note: Unpublished tests for differences between EQRs with juveniles included versus excluded within CSEMP data using analysis of variance (ANOVA) yielded probability >0.05.

2.4 Managing data standardisation

To successfully combine data from a variety of sources, the use of a taxonomically ordered database has proved essential (ordering of data, elimination of synonyms and so on). It also allowed the application of standardisation rules to create a standard data matrix.

In December 2010, the taxon list of the MBITT R&D database (Unicorn©) contained over ten thousand different taxa, with data relating to a range of WFD biological quality elements that may be recorded during invertebrate sampling (for example, fish, algae). Taxa can be categorised at differing levels of taxonomic hierarchy (Table 2.2).

Table 2.2 Total number of taxa within Unicorn© R&D database (December 2010) under the different taxonomic levels

Taxonomic level	Number of taxa in database
Kingdom	4
Phylum	47
Class	92
Order	326
Family	1,281
Genus	3,148
Species	5,873

Standardisation rules require application on a case-by-case basis to determine whether each taxa should be removed (non-invertebrate and/or planktonic), enumerated as 1 (colonial) or retained unaltered in accordance with the updated standardisation criteria (Table 2.1). Ideally, taxa would be flagged for standardisation at the species level. However, given the practical considerations in assigning standardisation rules to each of the approximately 5,900 species and that characteristics that dictate standardisation are likely to differ less within taxonomic groups at lower taxonomic levels, it was accepted that assigning standardisation rules could take place at a taxonomic level higher than species level. Standardisation at the family level (and above) provides a compromise of sufficient resolution without entailing excessive timescales to complete. Standardisation codes were assigned to each family by members of the MBITT based on a range of literature (for example, Hayward and Ryland 1995, MarLIN (http://www.marlin.ac.uk)) and taxonomic expertise (Appendix A).

As further habitats are included in WFD assessment and AQC methodologies evolve, rules for standardisation will need to be maintained and updated.

3 Development of the Infaunal Quality Index (IQI)

The WFD requires the ecological health of the biological quality elements to be expressed numerically as an Ecological Quality Ratio (EQR). This chapter describes how the WFD compliant metrics were selected and combined within the Infaunal Quality Index to fulfil this requirement.

3.1 Background

Benthic invertebrates have long been used for water quality assessments, with a range of methods having previously been developed to assess the health of benthic assemblages based on assemblage structure and function. However, when WFD implementation was first considered in 2002, there were no established, fully WFD compliant, assessment methods for benthic invertebrates in transitional and coastal waters. It therefore proved necessary for EU Member States to develop suitable assessment tools in order to fulfil their legal requirements and to provide effective management tools for the WFD.

In terms of the use of metrics as methods for measuring biological integrity, there has been a more recent tendency to move from single metric to multimetric assessment, as multimetric methods have proved more effective indicators of disturbance. A metric is a measure of the biota that changes in some predictable way with increased human influence (Barbour et al. 1995). The approach taken by MBITT was not to develop new metrics but rather to combine suitable existing metrics to establish a WFD compliant classification tool. Further discussion of the considered metrics and the multimetric approach is presented in the interim technical report (Prior et al. 2004).

Each WFD biological classification tool needs to assess the status of an individual biological quality element against a reference (background) level (COAST 2003). The actual components that need to be considered when assessing the quality element are described in the WFD normative definitions (WFD Annex V). For TraC benthic invertebrates, high ecological status (the upper proportion of which is considered as reference condition) is described as:

> 'The level of diversity and abundance of invertebrate taxa is within the range normally associated with undisturbed conditions. All disturbance-sensitive taxa associated with undisturbed conditions are present'.

To provide a WFD compliant assessment for the five ecological status classes (high, good, moderate, poor and bad), the benthic invertebrate assessment tool therefore needs to consider:

- diversity
- abundance
- disturbance sensitive taxa
- taxa indicative of pollution (bad to moderate definitions)

3.2 Requirements of the IQI

As well as fulfilling the specific requirements of the WFD normative definitions, in order to be an effective management tool it is recommended that a proposed classification scheme meets a range of additional criteria (Gibson et al. 2000). The assessment scheme should:

- demonstrate predictable change with increasing degrees of stress
- be specific to anthropogenic disturbance
- be sensitive to stress at low levels
- be applicable to a wide range of transitional and coastal waters
- be easily understood by non-specialists

As such it was necessary to ensure that the IQI was developed to deliver the functionality described below.

3.2.1 Ability to identify departure from biological reference conditions

The classification tool must be able to address the parameters defined in the normative definitions (diversity, abundance, disturbance sensitive taxa, taxa indicative of pollution – see Chapter 5) and use the characteristics to describe each of the five ecological status classes. The deviation from reference conditions to the class boundaries, as reflected in the classification index, must be quantifiable.

3.2.2 Responsive to anthropogenic pressures

Departures from reference conditions, reflected by the classification tool, must be demonstrated to link to anthropogenic pressures. Macrobenthic assemblages are exposed to varying degrees of natural stress depending on location (for example, salinity variations, aerial exposure), but any classification assessment must be able to distinguish an assemblage's response to anthropogenic pressures from those changes resulting from natural pressures.

It is widely accepted that natural stress is particularly elevated in transitional waters, resulting from highly variable physicochemical conditions such as salinity, dissolved oxygen, pH, hydrodynamics and temperature. The major challenge of identifying the symptoms of stress resulting from anthropogenic activities in such naturally stressed conditions has been termed the Estuarine Quality Paradox (Dauvin 2007, Elliott and Quintino 2007). When developing a benthic classification, it has to be recognised that many of the available measures of ecological health typically found under naturally stressed conditions display many of those characteristics associated with assemblages from areas suffering from anthropogenic stress (for example, the absence of full complement of k-strategists; Elliott and Quintino 2007). The underlying principle of the Estuarine Quality Paradox may be applied (albeit to a lesser degree) to other aspects of the physicochemical environment in both coastal and transitional waters, such as detecting the effects of anthropogenic disturbance under naturally anoxic conditions in sediments naturally dominated by fine particles for example, where >30% of sediments are <200 µm (Boaden and Seed 1988).

Due care needs to be taken to ensure that classification tools consider the variability of the biological assemblages in response to natural versus anthropogenic stress.

This can be approached by either quantifying and explaining the natural stress and subtracting this from the overall stress (thus isolating that from anthropogenic sources), or having an alternate set of methods which can detect anthropogenic stress but against a background of natural stress (Elliott and Quintino 2007). Without this recognition of the inherent stress of the system, any programme of measures initiated to drive environmental improvements, as a result of the classifications, may be fundamentally flawed. The approach proposed to quantify natural variability and incorporate its effects within the classification tool is described in Section 3.7.

3.2.3 Proportional in response to anthropogenic pressures

A monotonic (linear or curvilinear) correlation of the classification with anthropogenic pressures, where maximum and minimum classification score values represent the extremes of anthropogenic disturbance, is required. In such cases, the extent of the pressure can be estimated.

The use of metrics that do not have a monotonic response to increasing anthropogenic pressure may be unsuitable in reflecting a pressure gradient over the linear EQR scale as prescribed by the WFD. Metrics with a non-monotonic response to pressure will display a maximum or minimum value at a midpoint over a pressure gradient. Departure from this maxima/minima in either direction over the gradient (that is, whereby the pressure either increases or decreases) will result in the same response by the metric. For such metrics, a single metric score will therefore be indicative of pressure at two or more separate points along the gradient. This can be illustrated using the general model of change over a disturbance gradient where a single value for total abundance (X) is indicative of two separate points (A and B) over the disturbance gradient (Figure 3.1) which may equate to different WFD status classes. Note: this does not apply if the maximum or minimum occurs completely within either the bad or high status class, where reduction or increase in status is prevented by the lower and upper end of the metric scale respectively.

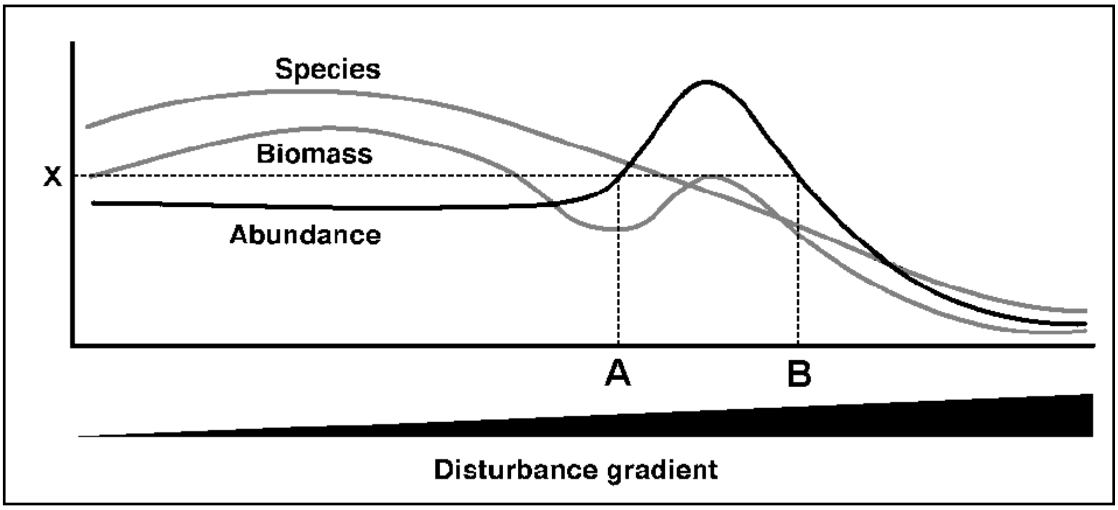


Figure 3.1 General model of changes in species, abundance and biomass along a gradient of increasing environmental disturbance (after Pearson and Rosenberg, 1978)

Note: Model annotated to illustrate the occurrence of abundance X at points A and B over the disturbance gradient.

3.2.4 Expressible quantitatively

WFD assessments are expressed as Ecological Quality Ratios. Annex V, section 1.4.1(ii), the WFD states that:

> 'To ensure comparability of such monitoring systems, the results of the systems operated by each Member State shall be expressed as ecological quality ratios for the purposes of classification of ecological status. These ratios shall represent the relationship between the values of the biological parameters observed for a given body of surface water and the values for these parameters in the reference conditions applicable to that body. The ratio shall be expressed as a numerical value between zero and one, with high ecological status represented by values close to one and bad ecological status by values close to zero'.

The assessment therefore needs to operate on a scale of zero to one: zero reflecting ecological quality under extreme anthropogenic disturbance and one representing ecological quality where anthropogenic disturbance is absent or negligible (Figure 3.2).

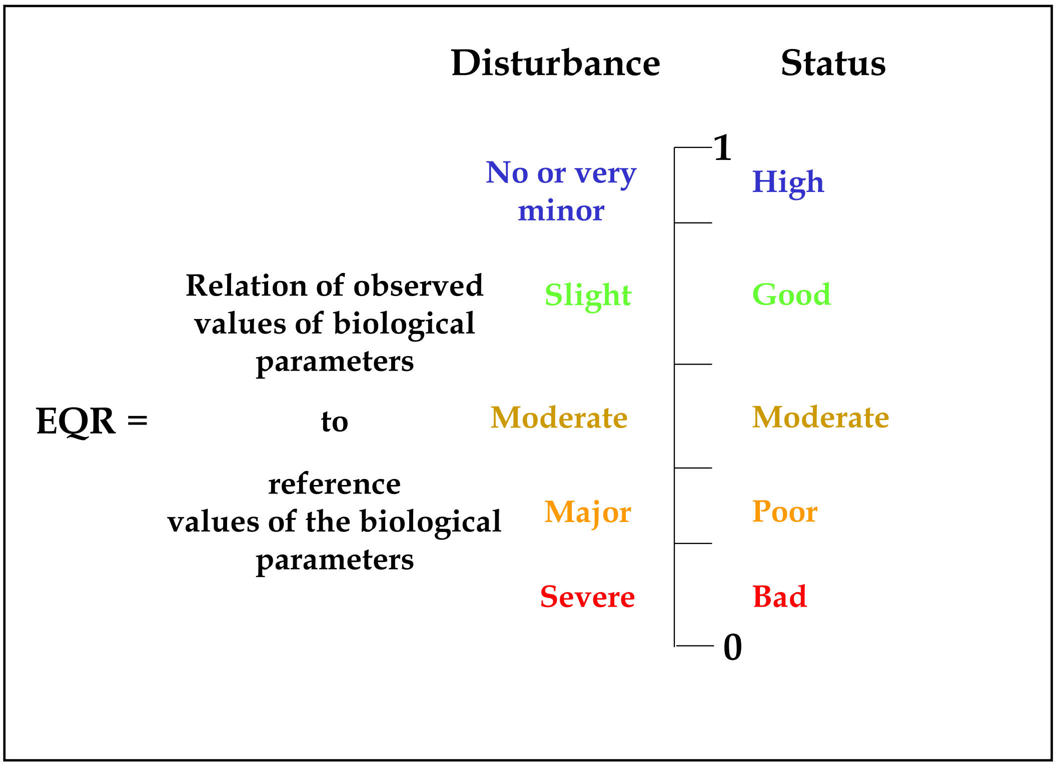


Figure 3.2 Suggested EQR according to WFD Annex V, section 1.4.1. Sizes of the bands may differ because the boundaries between classes must align with the normative definitions, not a simple percentage (all deviations are measured from the reference condition; COAST 2003)

3.2.5 Compatible with existing monitoring methodologies

The WFD specifies that the monitoring of parameters must conform to ISO/CEN standards[4] (or international or national equivalents where available) to ensure comparability with results across organisations and Member States. The compatibility of the classification method with common sample collection, processing and analysis methods also enables the development and testing of the classification methods on an extensive range of existing data. Many, if not all, of the WFD compliant biological metrics (see Section 3.3) have a degree of dependency on the methods by which a benthic sample is collected and processed, so common methods should be followed to ensure the metrics are derived consistently.

3.2.6 Independent from sample effort

The classification tool must be either independent of sampling effort, or adaptable for consistent results between different levels of sample effort. The classification tool should be developed so that changes in sample effort only serve to increase the confidence in assessment results and not the value of the EQR itself.

3.3 IQI development process

The IQI was developed as a classification index to enable the objective and quantitative estimation of the ecological status of a macrobenthic infaunal assemblage, integrating information from existing indicators and to be compliant with the normative definitions of the WFD.

Individual biological metrics have particular benefits and drawbacks (see Washington 1984), but a single metric used in isolation typically calculates only one measurable characteristic of an assemblage. This limits the suitability of any of the commonly used metrics to appropriately describe more than one of the WFD defined invertebrate characteristics.

A combination of metrics enables reference and degraded samples to be distinguished more effectively than one metric alone (Weisburg et al. 1997, Gibson et al. 2000). This is attributed to the staged response of benthos to stress, in which different metrics display greater response with different degrees of perturbation. As a result of the limitations of using single metrics and the benefits of using multiple metrics, the IQI followed a multimetric approach.

3.3.1 WFD compliant metrics

The effectiveness of the multimetric approach hinges on the performance of the underlying component biological metrics. To fulfil the requirements of the WFD, the metrics must be able to differentiate between the ecological status classes (high, good, moderate, poor and bad) in accordance with the indicators of biological health defined in the normative definitions (that is, diversity, abundance, pollution sensitive and pollution indicative taxa).

The suitability of a suite of benthic assemblage metrics that met those requirements was investigated at the start of the WFD tool development process (Prior et al. 2004).

[4] For benthic invertebrate sampling, ISO 16665:2005, *Water quality – Guidelines for quantitative sampling and sample processing of marine soft-bottom macrofauna*.

The structural and functional metrics considered are listed in Table 3.1. The strengths and weaknesses of the metrics are discussed in the interim technical report (Prior et al. 2004) and are not therefore reviewed in this report.

Table 3.1 Metrics considered in the development of the IQI [1]

Structural metrics	Functional metrics
Number of taxa (S)	Infaunal Trophic Index (ITI)
Abundance (N)	AZTI Marine Biotic Index (AMBI)
A/T (ratio of dominance: abundance/number of taxa)	
Shannon Weiner (H′)	
Pielou (J′)	
Fisher's α	
Brillouin (H)	
Simpsons (1-λ')	
Margalef (d)	
Average taxonomic distinctness (AvTD)	

Note: [1] While measures of biomass are widely agreed as being valid indicators of the health of macrobenthic assemblages (and can therefore provide useful measures of anthropogenic disturbance), metrics based on biomass were excluded from the IQI development as they are not a requirement of the WFD. Assessment of biomass may be included, if required, within WFD investigations.

3.3.2 Metric selection

The approach aimed to incorporate a combination of structural and functional elements within an index on the basis that it is likely to improve the ability of an assessment in determining if an area is anthropogenically affected (Elliott and Quintino 2007). Additionally, for considering the potential use of an index for classifying transitional waters, Elliott and Quintino (2007) also highlighted the deteriorating effectiveness of structural indices in habitats exposed to high degrees of natural stress where relatively few taxa exist and the potential increase in robustness of functional measures under such conditions.

For a multimetric to effectively identify changes to a benthic assemblage, the incorporated metrics should, in combination, explain the greatest degree of variability within of the data, where each metric should complement the other metrics, by explaining changes to the benthic assemblage that are not identified by the other metrics. It is not necessary to include metrics that correlate strongly, as these reflect similar variation within the assemblage and the inclusion of both is likely to provide little additional information on the ecological status of the assemblage over using only one of highly correlating metrics.

To aid in the initial selection of metrics for the WFD classification tool, principal components analysis (PCA) was performed on the considered metrics (Table 3.1) which had been calculated for benthic invertebrate data from a specific habitat type, thus lowering the effect of natural variability (see Prior et al. 2004). PCA provides a method by which to select those metrics that best describe the variation shown in the benthic invertebrate assemblage. The distribution of metrics within the PCA relates to their similarity with those metrics that cluster together describing similar variation within the data. Metrics that account for different parts of the variation were selected for incorporation into the IQI.

The PCA was initially carried out on metrics calculated for benthic invertebrate data collected from stable, depositional sediments under the UK's Clean Seas Environment Monitoring Programme (CSEMP). Abundance (N) and the Infaunal Trophic Index (ITI-UK) were separated from all the other univariate, diversity and functional metrics (Figure 3.3). When N and ITI-UK were excluded from the PCA, species richness (S) and AZTI Marine Biotic Index (AMBI) emerged to explain further the variation within the dataset (Figure 3.4). As a result, N, ITI-UK, S and AMBI were initially considered for inclusion within the WFD multimetric. Further analysis refined the selection to N, S, AMBI and Simpson's evenness, and adapted the form in which the metrics were included (see later sections).

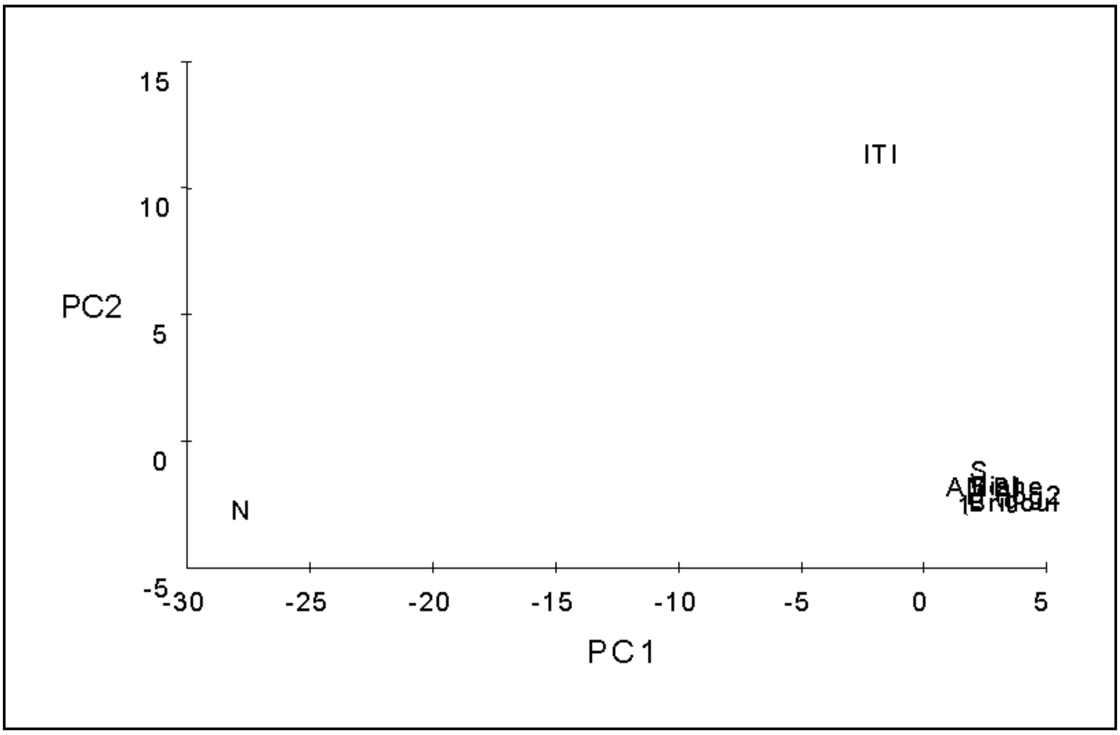


Figure 3.3 PCA of univariate, diversity and functional metrics calculated for CSEMP 0.1 m^2 Day grab data (>0.5 mm mesh, n = 95, Prior et al. 2004)

Notes: AMBI, AZTI Marine Biotic Index; ITI-UK, Infaunal Tropic Index UK; N, abundance; PC1, first principal component; PC2, second principal component; S, species richness.

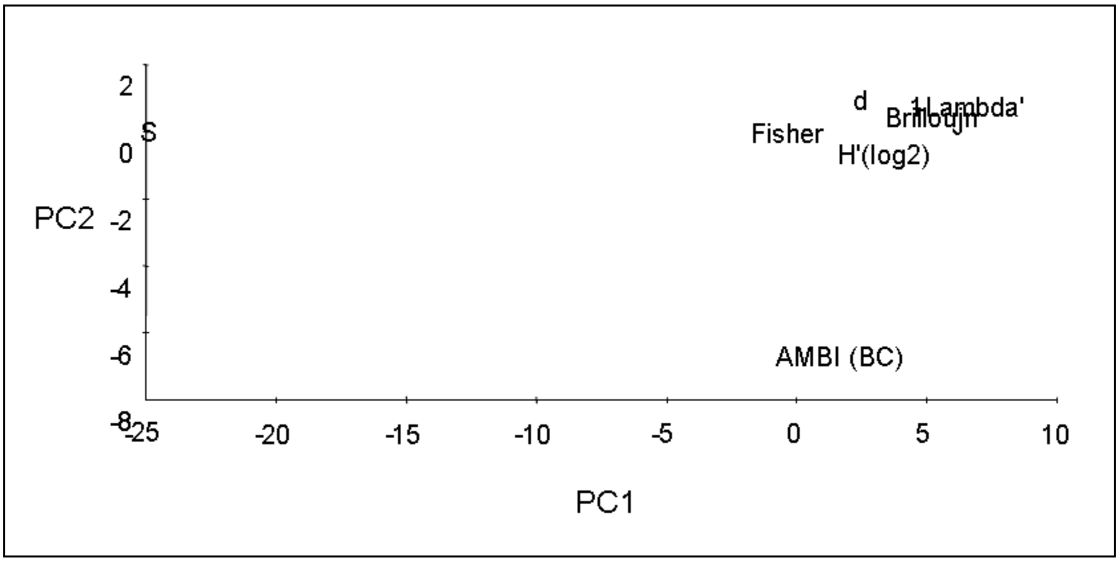


Figure 3.4 PCA of univariate, diversity and functional metrics calculated for CSEMP 0.1 m^2 Day grab data (>0.5 mm mesh, n = 95) with N and ITI-UK removed to expand clustered metrics within Figure 3.3 (Prior et al. 2004)

3.3.3 Data used for classification tool development

Once the initial metrics were selected using PCA, it was necessary to refine the selection by testing the response of the metrics against quantifiable pressure gradient data. As the proposed metrics are, to some extent, dependent on sample collection and processing methods, test data were all based on methods comparable to those adopted by UK agencies for national monitoring. National monitoring methods are outlined in the CSEMP monitoring manual (Cefas 2009)

Two key datasets were used in the early development stages of the IQI:

- the Garroch Head sewage sludge disposal ground (dataset courtesy of the Fisheries Research Services (FRS)
- the Boulby Coast Cleveland Potash mining discharge (dataset courtesy of AstraZeneca)

The resulting classification index was further tested against additional, independent, pressure datasets. Note that, prior to testing, some data standardisation was required to ensure that outcomes reflected environmental impact rather than differences in methodology. Chapter 2 describes the data pretreatment rules applied to the benthic invertebrate data.

Garroch Head sludge disposal ground dataset

The Garroch Head sludge disposal ground dataset provided a quantifiable organic enrichment gradient, with associated metals, for testing the behaviour of the IQI and its ability to ascribe change in biological assemblages to anthropogenic pressure.

Annual monitoring surveys of the site were undertaken from 1979 to 1997 to comply with consent conditions imposed by the licensing authority, Scottish Office Agriculture and Fisheries Department (SOAFD, formerly the Department of Agriculture and Fisheries for Scotland).

The Garroch Head Sludge disposal ground lies offshore of the Ayrshire coast beyond the mouth of the Clyde (Figure 3.5). The sludge discharge operated between 1904 and 1998, with discharges of between 1.576×10^6 and 1.740×10^6 tonnes of sewage sludge per year between 1987 and 1998. Discharging was undertaken by Strathclyde Regional Council (1979-1995) and the West of Scotland Water Authority (1996-1998) under licence by SOAFD. Disposal ceased in 1998 in order to meet the UK's obligations under the EU Urban Waste Water Treatment Directive (91/271/EEC).

Reports by the Scottish Marine Biological Association (surveys from 1979 to 1988) and SEAS Ltd (surveys from 1989 to 1997) on the annual monitoring surveys described the benthic communities as grossly distorted within 1.5 km of the centre of the disposal ground, with disturbance still discernible at 3 km from the centre. As such, the monitoring data provided a measurable impact gradient for use in developing the WFD classification index.

Data from 10 sites from 1979 to 1998 (inclusive) were used for the IQI development; consisting of 176 benthic infaunal samples, each supported by quantitative contaminant data. Macrobenthic sampling was undertaken between May and June of each year with samples taken at varying distances (up to approximately 1 km) from the disposal site. Single 0.1 m^2 Van Veen grab samples were taken at each sampling station between 1979 and 1985, with two 0.1 m^2 Van Veen grab samples being taken at each station from 1986 onwards. Macrobenthic samples were processed through a 1 mm sieve to retain the macrofauna. Analytical quality control was applied for the purpose of sorting efficiency and identification consistency.

Sediment chemistry data (acidity, redox potential, total carbon, total nitrogen, organochlorine compounds and metals (Cd, Cr, Cu, Hg, Ni, Pb, Zn, As, Co, Mn, Fe)) were taken at each station and qualitative sediment descriptions were recorded. Station data used for the IQI development had substratum described as either silt or silt/clay. All stations were fully marine (salinity of ~ 34) with depth ranging from 60 m to 180 m.

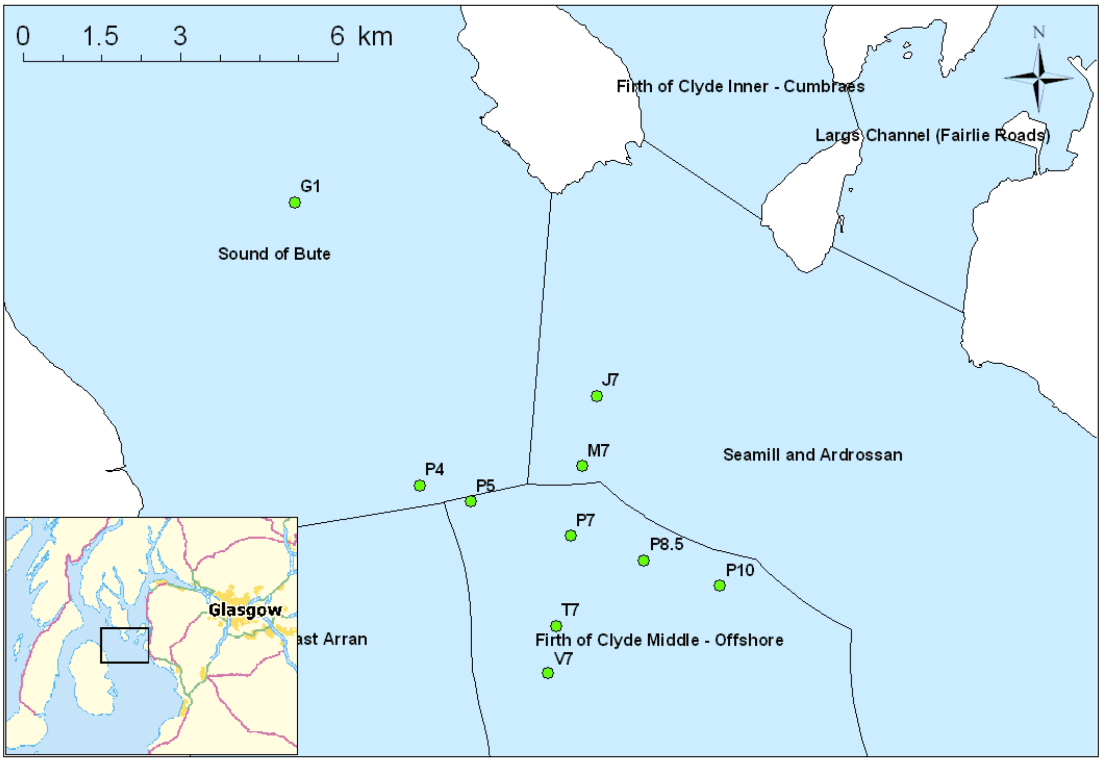


Figure 3.5 Benthic sampling stations surrounding the Garroch Head sewage sludge disposal ground

Note: Station P7 indicates the approximate location of the disposal ground, with the reference site located at station G1.

Cleveland Potash mine waste discharge dataset

The Cleveland Potash dataset also provided quantifiable pressure data (metals and sediment deposition) for testing the ability of the IQI to describe change in biological assemblages as a result of anthropogenic impact.

Waste products from the Cleveland Potash mine are disposed of (consented discharge) onto the Boulby Sandpatch on the North Yorkshire coast between the towns of Whitby and Saltburn-by-the-Sea, with reference sites at Runswick Bay, Sandsend and Skinningrove.

Although ecological monitoring has been conducted at Boulby since 1970, only data from 1998 to 2002 were used for development of the IQI. Annual surveys were carried out to assess the extent of accumulation of the waste material and to assess ecological impact to benthic macroinvertebrate fauna; the mining process discharged in the order of between 120,000 and 170,000 tonnes of insoluble waste annually (Brown and Shillabeer 1997). Monitoring was coordinated by the Boulby Monitoring Working Group, representing Cleveland Potash, the Environment Agency, Cefas and AstraZeneca's Brixham Environmental Laboratory.

Results of the macrobenthic analysis, sediment chemistry and particle size analysis (PSA) were provided in a series of reports, *Marine Monitoring for Cleveland Potash Ltd* (AstraZeneca 1999-2003). Brown and Shillabeer (1997) concluded that the accumulation of insoluble material around the discharge correlated to reduced diversity and abundance in the local infaunal community around the discharge. Sedimentary concentrations of the metals/metalloids boron, copper and mercury were elevated immediately around the discharge, but bio-accumulation of these metals at the discharge site was indistinguishable above that at the reference stations.

For the IQI development, subtidal data from 1998 to 2002 (inclusive) were used. At Boulby, samples were taken at varying distances up to approximately 1km from the two discharge sites (Figure 3.6). Triplicate Day grab samples were taken at the majority of reference sites and at a subset of Boulby stations. Single samples were taken at all other stations. The dataset from the 21 sites over study period consists of 213 benthic infaunal samples, with 105 associated physicochemical samples. Macroinvertebrate samples were taken using a 0.1 m^2 Day grab during September of each year and processed through a 1 mm sieve. Benthic fauna were identified to the lowest practicable taxonomic level, with abundance determined according to Brixham Environmental Laboratory's Standard Operating Procedure (SOP NS015). Analytical quality control was applied for the purpose of sorting efficiency and identification consistency. Sediment samples for particle size and metals analysis (B, Cd, Cr, Cu, Hg, Pb, Ni, Zn) were taken from a 0.1 m^2 Day grab with a 20 mm diameter sediment core. Sediment particle size distribution was determined for 12 size bands between <4 μm and 64 mm, and the percentage of the silt/clay fraction (SOP NS024).

The sampling locations for the Boulby Coast were typically dominated by medium to fine sands (63–500 μm). The exception was within the close vicinity of the discharge positions where the substratum became dominated by silt/clay (<63 μm), considered to be a result of the discharged mining waste (Brown and Shillabeer 1997).

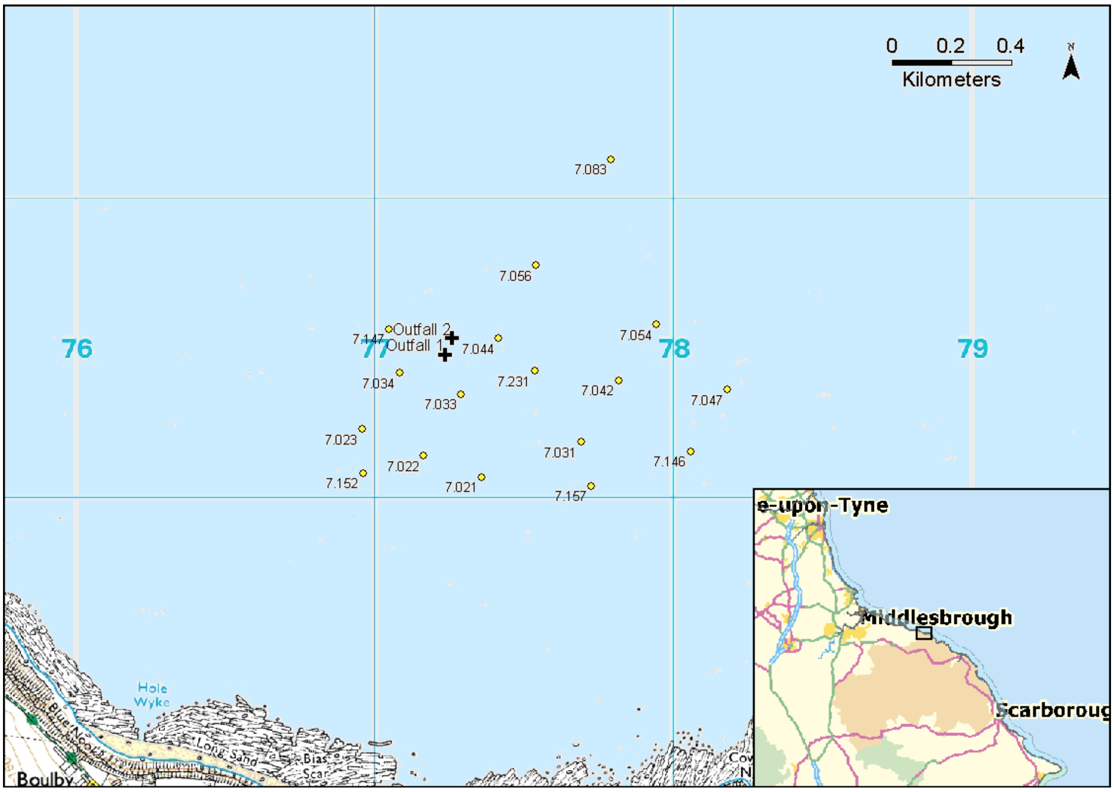


Figure 3.6 Benthic stations surrounding the Boulby Coast Cleveland Potash mining discharge locations

3.4 Formulating the IQI

The IQI was developed in a series of phases, with modifications occurring as understanding of the WFD requirements advanced and further WFD compliant data became available. The development was an iterative process where stepwise review and refinement of the classification tool occurred. Each version was tested against defined datasets, incorporating expert judgement, and revised where necessary (Figure 3.7).

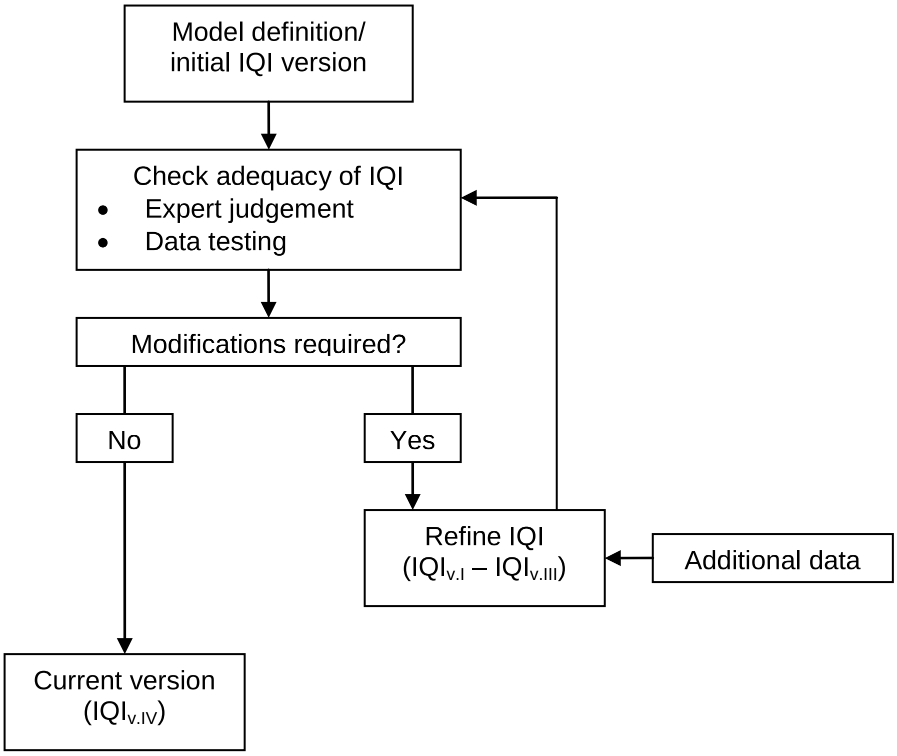


Figure 3.7 Generic process behind the development and refinement of the IQI

Four distinct revisions of the IQI have so far been undertaken ($IQI_{v.I}$ to $IQI_{v.IV}$). This report documents how $IQI_{v.IV}$, which was used in the UK's first RBMPs in 2009 to describe the status of the benthic invertebrate community, was formulated. The development of each revision of the IQI is described in Sections 3.5 to 3.8, and version-specific process charts (Figure 3.8 and Figure 3.9).

Although the underlying principles behind the adaptation of all versions are outlined in this report, full details of the development of the intermediate versions of $IQI_{v.I}$ to $IQI_{v.III}$ are not addressed unless relevant to the development of $IQI_{v.IV}$.

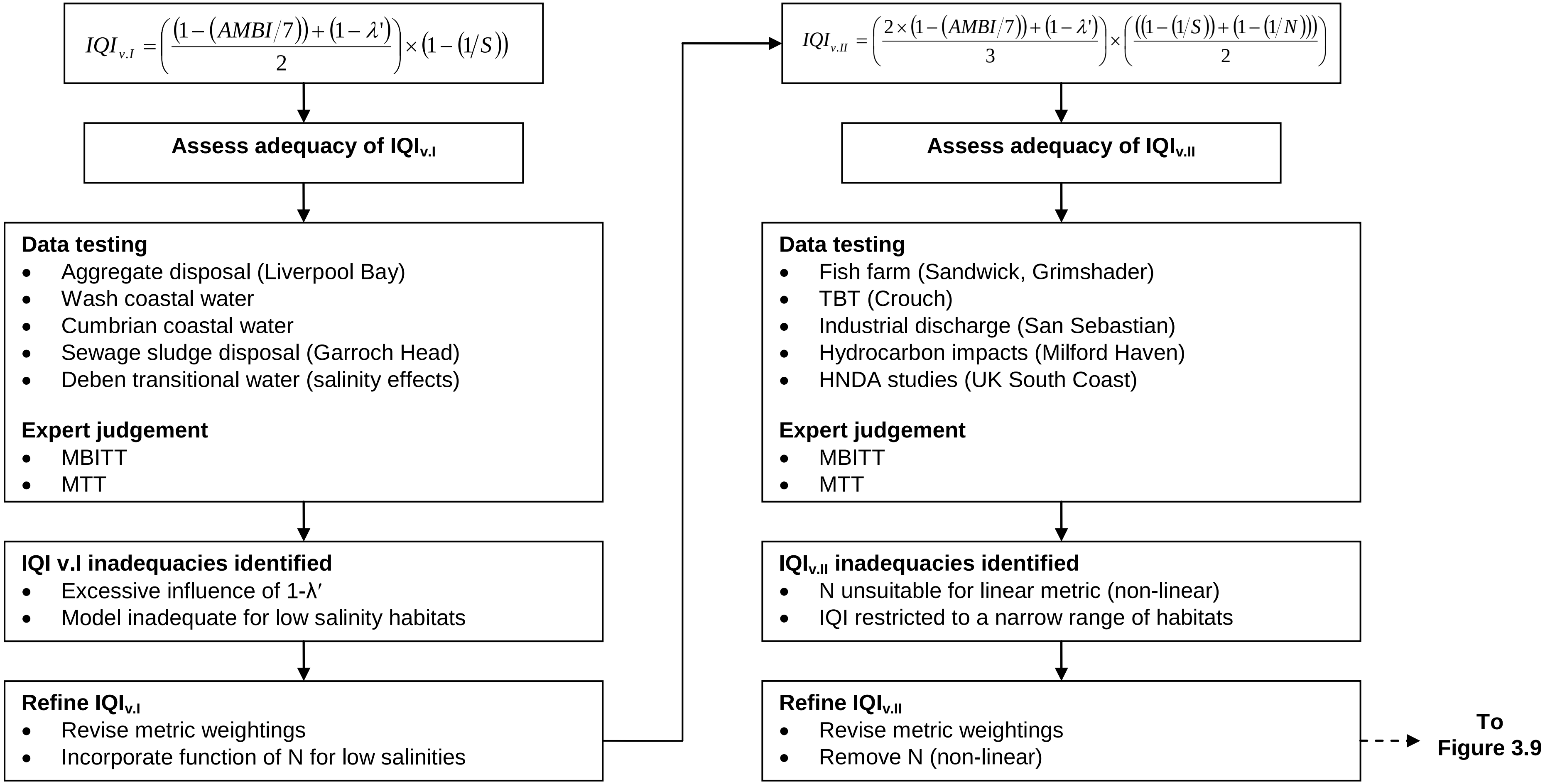


Figure 3.8 Process chart for the development of $IQI_{v.I}$ and $IQI_{v.II}$

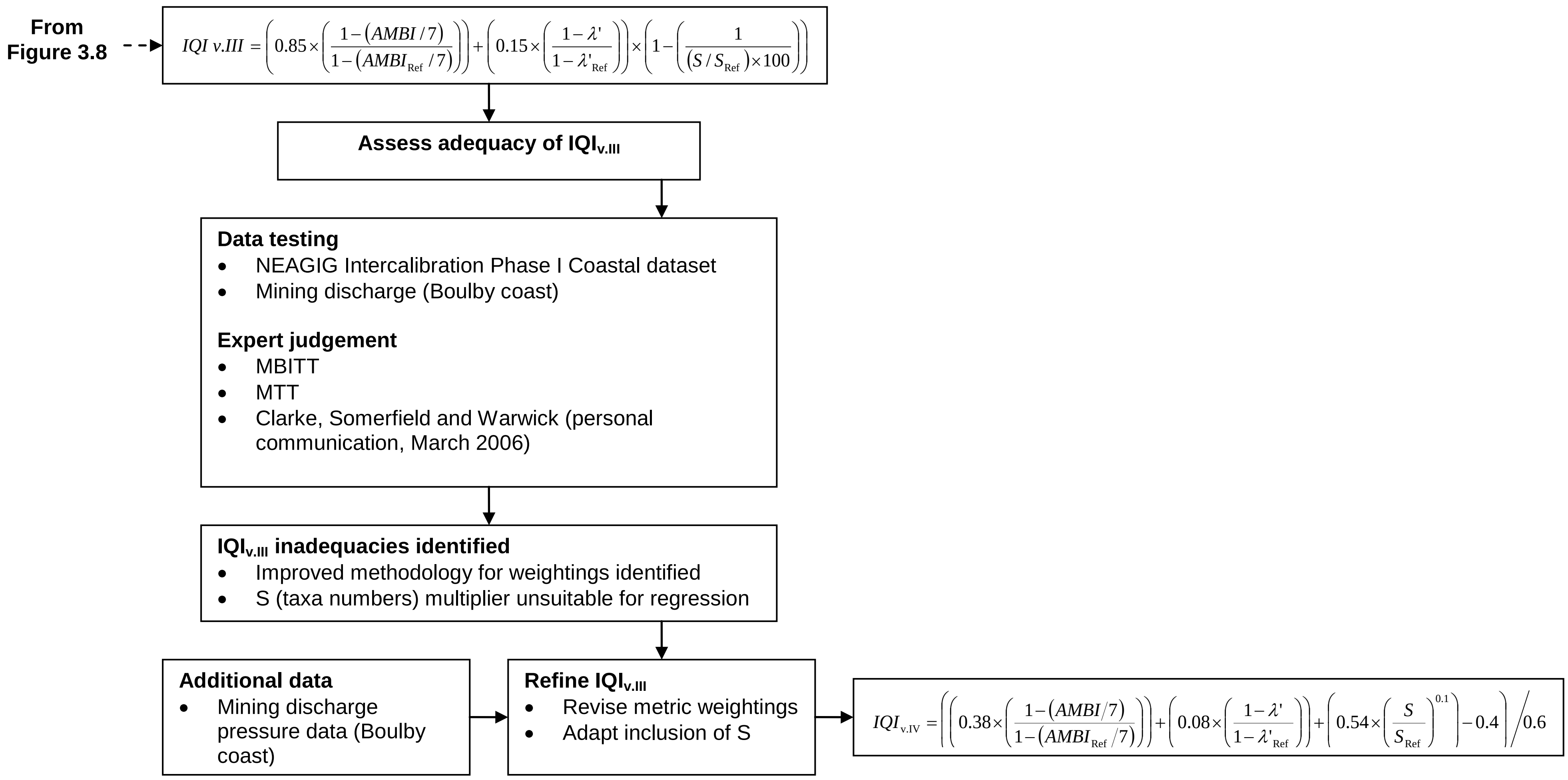


Figure 3.9 Process chart for the development of $IQI_{v.III}$ and $IQI_{v.IV}$

3.5 IQI version I (2003-2004)

$IQI_{v.I}$ was formulated as:

$$IQI_{v.I} = \left(\frac{(1-(AMBI/7)) + (1-\lambda')}{2} \right) \times (1-(1/S))$$

Equation 3.1

The main considerations were as described below.

3.5.1 Selection of suitable metrics

The metric selection process (PCA) identified ITI-UK, abundance (N), AMBI, Simpson's evenness (in the form 1-λ′) and taxa number (S) for initial incorporation within the IQI. Further consideration (expert judgement) refined the selection considering the scope for potential use (AMBI) and level of inherent variability seen in the abundance data (N).

3.5.2 The form in which to incorporate the metrics

The EQR operates between zero (severe impact) and one (minimal disturbance). The AMBI operates over the scale of seven (azoic) to zero (100% ecological group 1 (EGI) taxa) (see Borja et al. 2000). To align with the requirements of the WFD, the IQI required the AMBI to be normalised and inverted to operate between zero (azoic) and one (100% EGI taxa). The AMBI was normalised and inverted to the EQR compatible scale as 1-(AMBI/7).

Simpson's evenness (in the form 1-λ′) operates over the scale of zero (assemblage dominated by a single taxon) to one (individuals evenly distributed among all taxa), which corresponded to the EQR scale of zero to one respectively without needing transformation.

1-λ′ and 1-(AMBI/7) operate in a similar manner over the zero to one scale which allowed the two metrics to be averaged to form the basis of the $IQI_{v.I}$ (Equation 3.2).

$$\left(\frac{(1-(AMBI/7)) + (1-\lambda')}{2} \right)$$

Equation 3.2

Taxa number (S) operates over a theoretical scale of zero (azoic) to infinity. S was modified to its reciprocal (that is, 1/S) to operate over a scale of one (S = 1) to zero (S = ∞), and inverted in the form 1-(1/S) to operate where zero to one corresponded to states of severe impact and minimal disturbance respectively. This modification was based on assessment of the Garroch Head impact gradient. Over this impact gradient, S displayed a curvilinear reduction from low to high concentration of contaminant (Cu) (Figure 3.10). At low levels of contaminants, S appears to vary between ~15 to ~65 taxa per 0.1 m^2. As a result, the inclusion of S within the assessment should only begin to reflect an anthropogenically impacted state where taxa numbers reduced below ~15 per 0.1 m^2. Above ~15 taxa, the balance of ecological sensitivity groups (AMBI) and evenness were considered to be the more appropriate drivers of the assessment. $IQI_{v.I}$ was therefore developed to incorporate taxa number whereby variation of S above 15 had negligible effect on ecological status. Alternative methods of the inclusion of S within the $IQI_{v.I}$ were explored. The relationship between S and 1-(1/S) is illustrated in Figure 3.11.

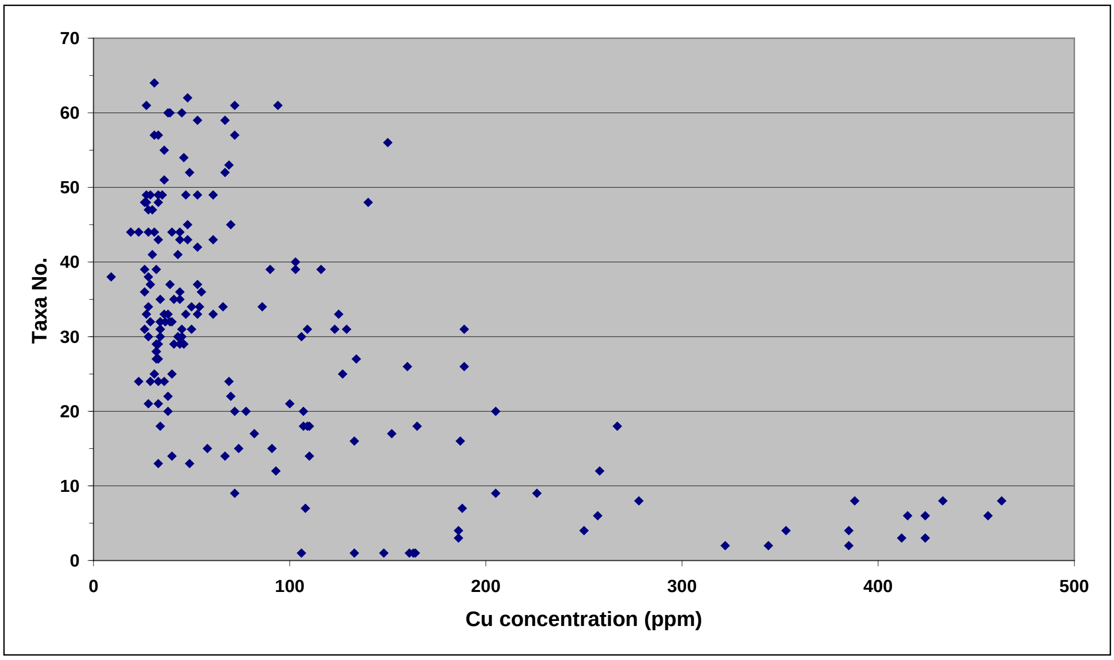


Figure 3.10 Taxa number (S) per 0.1 m^2 with Cu concentration (ppm) within the Garroch Head sewage sludge disposal ground data

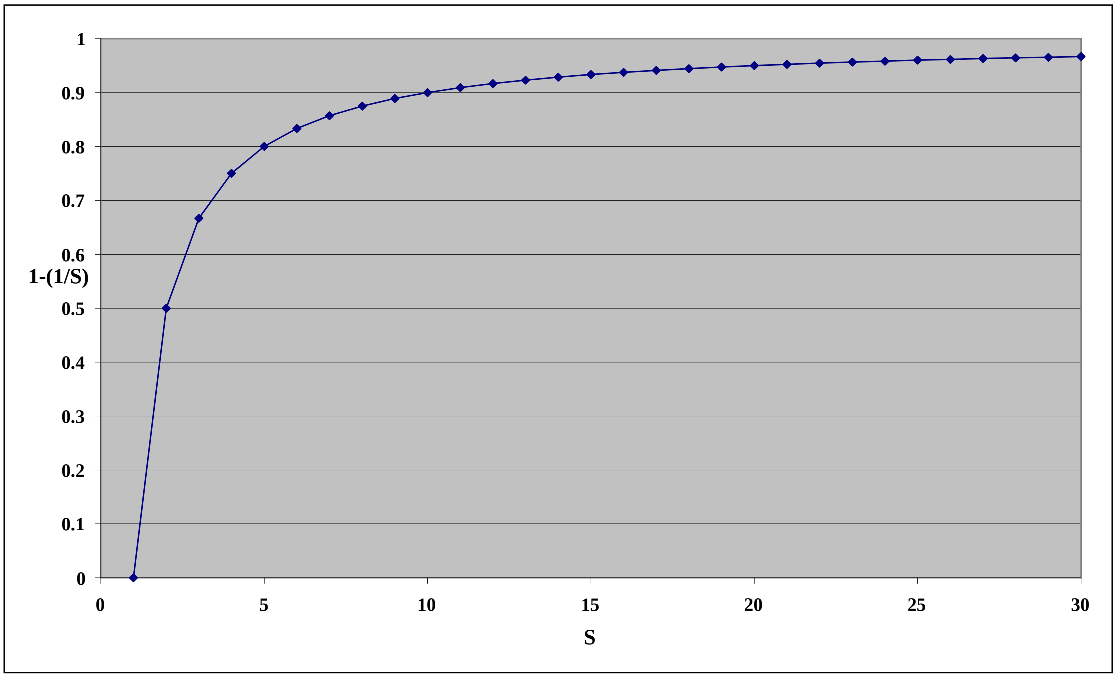


Figure 3.11 Relationship between S and 1-(1/S)

So that the effect of changing values of S on the multimetric score is negligible until <~15 taxa are present, S (in the form 1-(1/S)) was incorporated into the multimetric as a multiplier of the 1-(AMBI/7) and 1-λ′ components (Equation 3.1).

At this stage of development, the absolute values of S observed within the Garroch Head pressure gradient data were considered relevant only in the context of the habitat, pressure and methodologies specific to the dataset. It was presumed that:

- changes in S for different habitats (for example, low salinity, coarse sediments), pressures (for example, aggregate extraction) and methodologies (for example, 0.5 mm sieve mesh) would operate over different scales

- continued trial and development of the IQI was essential for it to apply to changes in such conditions

3.6 IQI version II (2004 to November 2005)

The $IQI_{v.II}$ was formulated as follows:

$$IQI_{v.II} = \left(\frac{2 \times (1 - (AMBI/7)) + (1 - \lambda')}{3} \right) \times \left(\frac{((1 - (1/S)) + (1 - (1/N)))}{2} \right) \qquad \text{Equation 3.3}$$

Review and testing necessitated the following amendments to the $IQI_{v.I}$:

3.6.1 Revision of metric weightings

Applying equal weighting to 1-(AMBI/7) and 1-λ′ in $IQI_{v.I}$ made the unsubstantiated assumption that the metrics were equally effective indicators of anthropogenic pressure. To avoid this assumption, different weightings of AMBI and 1-λ′ were tested to identify the weighting of the two metrics that provided the highest correlation to anthropogenic pressure. These weightings were applied to 1-λ′ and 1-(AMBI/7) for inclusion within the IQI as follows:

$$\frac{A \times (1 - (AMBI/7)) + B \times (1 - \lambda')}{A + B} \qquad \text{Equation 3.4}$$

where:

A = weight of 1-(AMBI/7)

B = weight of 1-λ′

To identify suitable values of A and B, Equation 3.4 was tested against the Garroch Head dataset (Section 3.3). Values for A and B were adjusted between 0 and 1. To ensure the sum of 1-(AMBI/7) and 1-λ′ operated between 0 to 1, the sum of the weightings A and B was fixed at 1 (for example, where A = 0.3, B = 0.7). Spearman rank correlation (ρ_w) was calculated between the metrics with different test weightings and normalised copper, the contaminant with the strongest correlation to the variability in the abundance data; the results are shown in Figure 3.12.

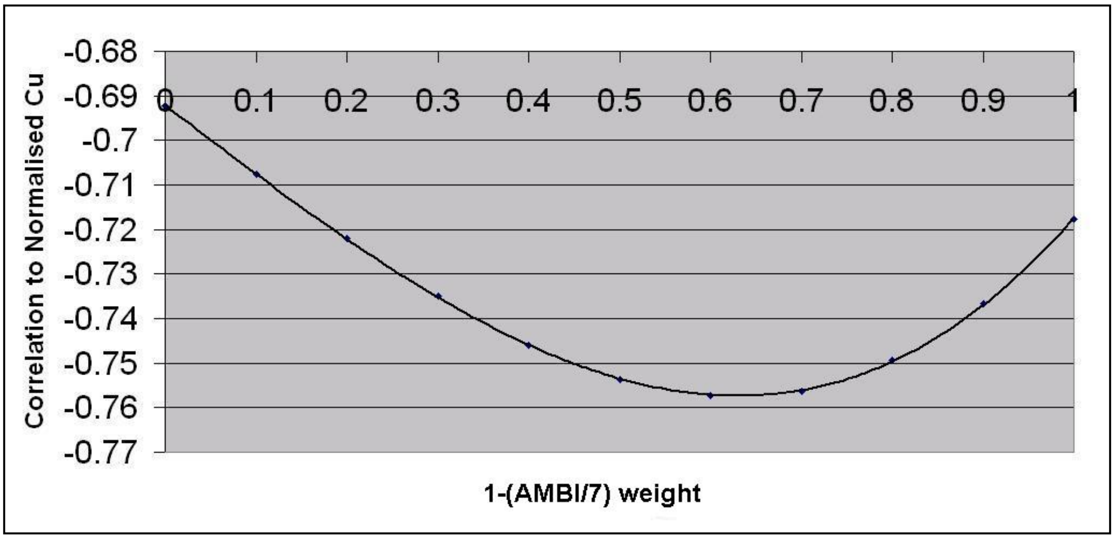


Figure 3.12 Spearman rank correlation (ρ_w) between normalised Cu (log (x+1)) and weighting of 1-(AMBI/7) in $IQI_{v.II}$ for the Garroch Head sewage sludge disposal ground data

The strongest correlation between Cu concentration and 1-(AMBI/7) weighting occurred where the 1-(AMBI/7) weight was ~0.64 (ρ_w = -0.757). The corresponding 1-λ′ weight was therefore 0.36 (total weight of both metrics = 1). The 1-(AMBI/7):1-λ′ weightings of 0.64:0.36 were approximated as a 2:1 weighting within $IQI_{v.II}$ (Equation 3.5).

$$\left(\frac{2\times\left(1-\left(AMBI/7\right)\right)+\left(1-\lambda'\right)}{3}\right) \qquad \text{Equation 3.5}$$

3.6.2 Adaptation for low salinity habitats

Testing the $IQI_{v.I}$ using data from low salinity locations (for example, Environment Agency Deben quinquennial macrobenthic survey 2002) generated sample statuses where very low numbers of taxa were present. This was inconsistent with what was expected given an absence of any considerable anthropogenic pressures.

Samples containing a single taxon were consistently assigned bad status, irrespective of taxon sensitivity and realistic ecological status as anticipated by expert judgement. The species richness component (1-1/S), used as a multiplier, returned EQR values of zero for samples populated by a single taxon, regardless of the abundance. Feedback from the wider project group suggested that the IQI should be capable of differentiating between samples with high and low abundance of single taxon. To accommodate this requirement, total sample abundance (N) was incorporated into the $IQI_{v.II}$ as a multiplier in the form 1-(1/N) (Equation 3.6).

$$\left(\frac{\left(\left(1-\left(1/S\right)\right)+\left(1-\left(1/N\right)\right)\right)}{2}\right) \qquad \text{Equation 3.6}$$

Therefore, the taxon number (S) and abundance (N) multiplier component is low only where both S and N are low. This results in an increase in EQR where one and/or the other values of S or N increase.

3.7 IQI version III (November 2005 to March 2006)

The $IQI_{v.III}$ was formulated as follows:

$$IQI_{v.III} = \left(0.85 \times \left(\frac{1-(AMBI/7)}{1-(AMBI_{\mathrm{Ref}}/7)}\right)\right) + \left(0.15 \times \left(\frac{1-\lambda'}{1-\lambda'_{\mathrm{Ref}}}\right)\right) \times \left(1 - \left(\frac{1}{(S/S_{\mathrm{Ref}}) \times 100}\right)\right)$$

Equation 3.7

Review and testing led to the following amendments to $IQI_{v.II}$.

3.7.1 Removal of total abundance (N) from the IQI

As illustrated by the Pearson–Rosenberg model (Pearson and Rosenberg 1978), total abundance (N) can be a useful indicator of anthropogenic disturbance. For certain pressures (for example, organic enrichment), moderate levels of perturbation result in elevated total abundance as populations of opportunistic taxa (r-strategists) increase. However, the relationship between total abundance and anthropogenic disturbance in such instances is not linear; as levels of perturbation increase from moderate to high, the tolerance thresholds of these opportunists is exceeded and total abundance then declines until a location becomes azoic.

Metrics displaying such nonlinear relationships with anthropogenic pressure are not suitable for inclusion in classification tools. Transformation of total abundance would be ineffective in providing a solution to this issue and so N was removed from the IQI.

3.7.2 Inclusion of metric-specific reference conditions

$IQI_{v.I}$ and $IQI_{v.II}$ were developed on the assessment of benthic data from a limited range of environmental conditions (fully saline, sublittoral sands and muddy sands and sublittoral muds as defined by EUNIS habitats A5.2 and A5.3[5]). EUNIS habitats are further addressed in Section 4.5.1. Development was also specific for sample collection and processing methods (0.1 m^2 grab, 1 mm sieve mesh). It can be demonstrated that the metrics tested and selected for use within the IQI are influenced by environmental conditions (habitat) and sampling methodology. For example, the number of marine taxa for a fixed sampling method shows a correlation with average salinity (Figure 3.13).

[5] Classified as A4.2 and A4.3 prior to the revision of the EUNIS classification system in 2004. More information on the use of EUNIS in the development of the IQI can be found in Prior et al. (2004).

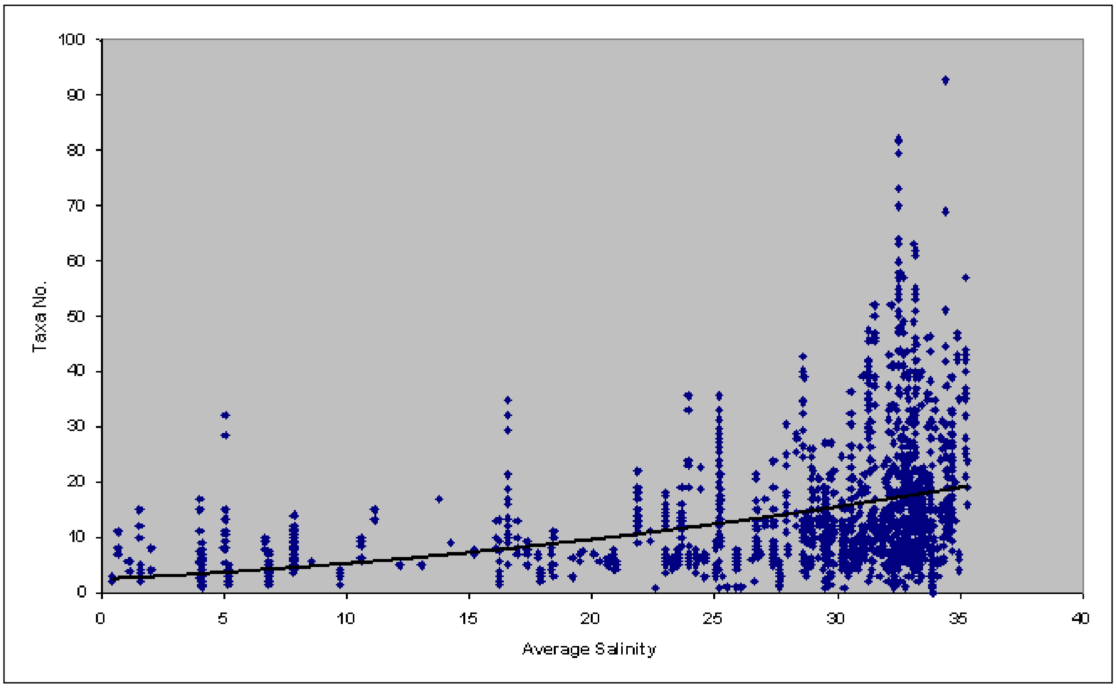


Figure 3.13 Taxa number (truncated) versus average salinity at station (taxa number corrected to sample area and mesh size)

Notes: Benthic data from Environment Agency R&D database 2001-2007.
Salinity data from Environment Agency database 1997-2007.

In order to apply the IQI to data collected from a broad range of habitats and a selection of sampling methods, changes in metric values as a response to natural stressors and differing sample collection and analysis methods must be acknowledged. That is to say, the IQI needs to be able to detect the extent of deviation from reference conditions once the effect of the natural pressures has been accounted for while avoiding the influence of sample method.

To discriminate between anthropogenic and natural stressors, the IQI needed to incorporate a mechanism to acknowledge the influence of natural pressures. Without this, the assumptions made in assessing an assemblage exposed to low natural stress (for example, fully marine subtidal muds) are the same as for communities exposed to high levels of natural stress (for example, estuarine, intertidal sediments).

To establish ecological class status, WFD assessments relate sample macrobenthic assemblages to reference macrobenthic assemblages. Acknowledgement of the different natural hydromorphological and physicochemical factors are considered within the IQI by reference to the habitat that the macrobenthic assemblage inhabits.[6]

Two options were explored with respect to the influence of habitat variation and how to incorporate suitable reference conditions:

- to adapt the ecological class status boundaries on a habitat by habitat basis

[6] It is acknowledged that the term 'habitat' is a superimposition of artificial boundaries used to separate and categorise the surroundings in which an assemblage resides according to hydromorphological, physicochemical and biological parameters which in reality operate over a continuum. The advantages and disadvantages of the categorisation into habitats are addressed further in Chapter 4 on the setting of reference conditions.

- to incorporate habitat specific reference values within the IQI

For the first option (adaptation of the boundaries) class boundaries would need to be set for a given habitat and applied following the EQR calculation. By deriving the extent to which changes in these habitats affects the classification tool, different class status boundaries could be applied accordingly, that is, boundaries would be lowered where the natural hydromorphological and physicochemical regime result in suppressed IQI values, and vice versa where elevated IQI values are a result of natural conditions. By incorporating habitat and method specific reference condition values into the IQI (second option), the resulting EQRs would be a direct expression of the ecological health of the assemblage relative to an appropriate reference condition. Status boundaries would remain static ensuring that the extent of departure from reference conditions represented by each class status was directly proportional for all habitats and sample methods.

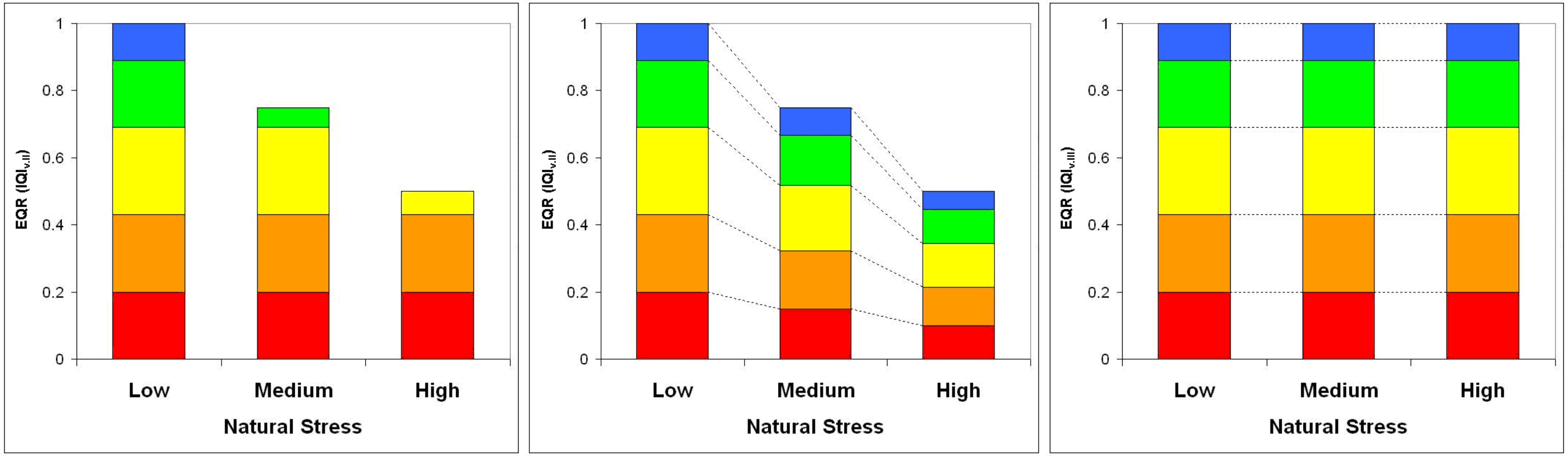
$IQI_{v.III}$ was developed following the second option, incorporating reference condition values within the multimetric. This was achieved by dividing the observed metrics by their 'expected' reference values for the relevant habitat and sample methods. Each metric in effect expresses the observed metric value as a ratio of what is expected under appropriate reference conditions. The reference values of the component metrics (AMBI, 1-λ′ and S) are labelled within the IQI as $_{Ref}$ values (that is, $AMBI_{Ref}$, $1-\lambda'_{Ref}$ and S_{Ref}).

Incorporating reference condition values as described above has the advantages outlined below.

Static boundaries for all habitats and methods

Class status boundaries relate to the extent of departure from reference conditions as described within the WFD normative definitions (class boundaries are addressed in detail in Chapter 5). The implications of incorporating reference conditions on the ecological status boundaries can be observed in Figure 3.14. Figure 3.14(a) illustrates how medium to high levels of natural stress may limit the typical upper EQRs if changes in reference conditions were not accounted for. Without incorporating reference condition values into the IQI as described above, the ecological status boundaries would need to be revised for habitats exposed to different levels of natural stress as illustrated in Figure 3.14(b). Under such circumstances, separate ecological status boundaries would be required for all combinations of habitat and sampling collection and processing methods. The need to modify class status boundaries according to habitat would increase the complexity of communicating status assessment, establishing environmental targets using the assessments and the boundary setting process through the NEAGIG intercalibration process.

Incorporating reference conditions within the IQI and adapting them to accommodate the effect of changing habitats and sample methods on the metrics, the scale of the IQI remains between 0 and 1 under all conditions. As points on the EQR scale represent a fixed relative extent of departure from reference conditions, this enables class status boundaries to be fixed for all habitats and methods (Figure 3.14(c)).



(a) (b) (c)

Figure 3.14 EQRs and corresponding ecological status derived for reference conditions exposed to low, medium and high levels of natural pressure for: (a) $IQI_{v.II}$ using fixed status boundaries; (b) $IQI_{v.II}$ adjusting boundaries according to extent of natural stress; and (c) $IQI_{v.III}$ adjusting reference conditions with fixed boundaries

Comparability between EQR values

By expressing each metric within the IQI as a ratio (habitat and method specific), the resulting EQR represents a ratio that has reduced dependency on the natural environmental conditions. This ensures clear comparability between EQRs, that is, EQR values relate to a specific ecological class (high to bad), regardless of habitat sampled or sampling methods used.

Reduced between sample variability (systematic bias)

The influence of changing habitat on the IQI metric values can be described as bias (or systematic variability). To illustrate, in the absence of changing pressures within a transitional water body, sample EQR values are likely to differ depending on the degree of natural pressure such as salinity. Without factoring this aspect into EQR calculations, this bias will serve to increase the variability between a set of samples within for example, a water body assessment. Accommodating this systematic variability in the IQI serves to reduce variability between EQR values from different habitats and, if based on data from multiple habitats, may serve to reduce the overall variability, thus allowing for increased statistical confidence in assessment results. The benefit of this from a management perspective is that it provides the potential to base monitoring programmes on fewer samples (and therefore lower cost) to attain a given degree of statistical confidence. EQR variability is addressed further in Chapter 6.

3.7.3 Adaptation of S

$IQI_{v.I}$ and $IQI_{v.II}$ included S as an absolute value in the form 1-(1/S). This required modifying so that S was represented as a relative value in terms of S_{Ref} in $IQI_{v.III}$. The approach adopted was therefore to retain a function of S as a multiplier, but to represent it as a percentage of S_{Ref} in the $IQI_{v.III}$:

$$1-\left(\frac{1}{(S/S_{\mathrm{Ref}})\times 100}\right)$$

Equation 3.8

3.7.4 Revision of 1-(AMBI/7) and 1-λ′ weightings

The weightings of AMBI and 1-λ′ in $IQI_{v.II}$ were established according to the correlation of different weightings of the two metrics prior to the addition of S into the analysis. It was subsequently considered that this approach might result in a suboptimal correlation between the IQI and the pressure data, as the inclusion of S may affect the extent to which the balance of AMBI and 1-λ′ correlate to pressure.

For development of $IQI_{v.III}$, the weightings of AMBI and 1-λ′ (or $AMBI_{Ref}$ and $1\text{-}\lambda'_{Ref}$) were therefore recalculated to include the influence of S (as expressed in Equation 3.8). As with the development of $IQI_{v.II}$, the weightings of $AMBI_{IQI}$ and $1\text{-}\lambda'_{IQI}$ that produced the maximum spearman rank correlation (ρ_w) between $IQI_{v.III}$ and pressure data (that is, normalised Cu) were adopted. These were $AMBI_{IQI}$ of 0.85 (ρ_w = -0.816) and $1\text{-}\lambda'_{IQI}$ of 0.15 (total weight of two metrics to equal 1, Figure 3.15). $IQI_{v.III}$ was therefore formulated as per Equation 3.7.

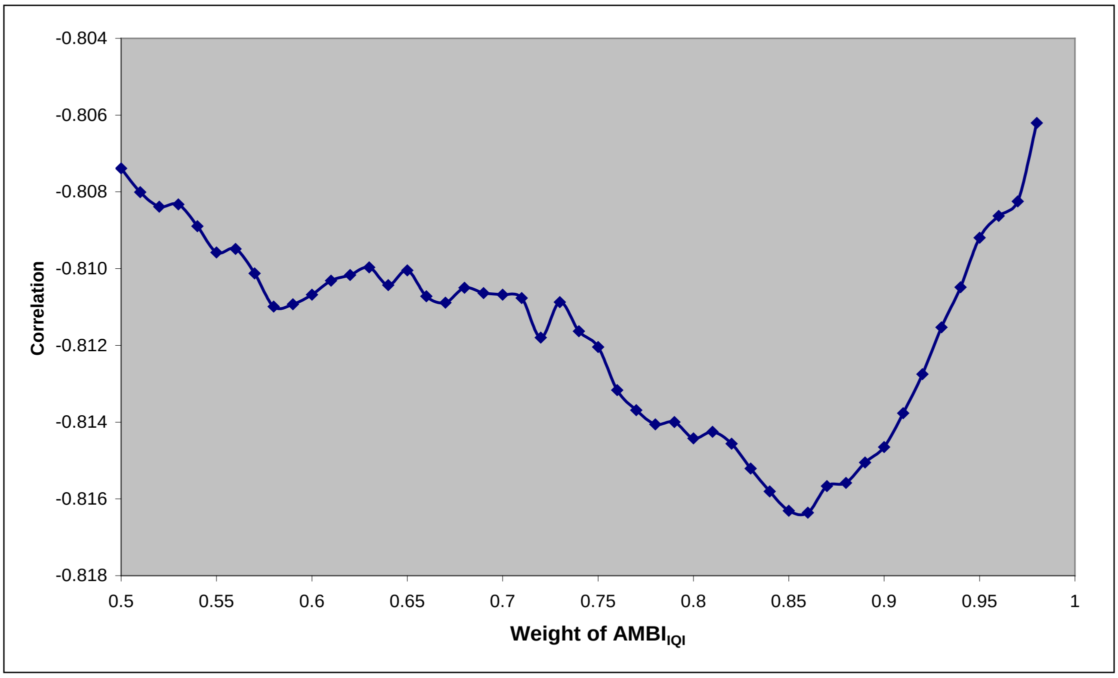


Figure 3.15 Spearman rank correlation (ρw) between normalised Cu (log(x+1)) concentration and weighting of AMBIIQI in $IQI_{v.III}$ for the Garroch Head sewage sludge disposal ground data

3.8 IQI version IV (March 2006 to May 2011)

The $IQI_{v.IV}$ is formulated as follows:

$$IQI_{v.IV} = \left(\left(0.38 \times \left(\frac{1-(AMBI/7)}{1-(AMBI_{\mathrm{Ref}}/7)} \right) \right) + \left(0.08 \times \left(\frac{1-\lambda'}{1-\lambda'_{\mathrm{Ref}}} \right) \right) + \left(0.54 \times \left(\frac{S}{S_{\mathrm{Ref}}} \right)^{0.1} \right) - 0.4 \right) \Big/ 0.6$$

Equation 3.9

The expression of the observed metrics divided by their reference condition values components from Equation 3.9 are referred to subsequently in this chapter as $AMBI_{IQI}$, $1\text{-}\lambda'_{IQI}$ and S_{IQI}. These can be substituted in Equation 3.9 to give:

$$IQI_{v.IV} = \left(\left(0.38 \times \left(\left(AMBI_{IQI} \right) \right) \right) + \left(0.08 \times \left(1 - \lambda'_{IQI} \right) \right) + \left(0.54 \times \left(S_{IQI} \right)^{0.1} \right) - 0.4 \right) / 0.6$$

Equation 3.10

Review and testing necessitated the following amendments to $IQI_{v.III}$:

- application of regression analysis to establish appropriate weightings of the IQI component metrics
- adaptation of the expression of S/S_{Ref}

The $IQI_{v.IV}$ development process is summarised in Figure 3.16 and Figure 3.17.

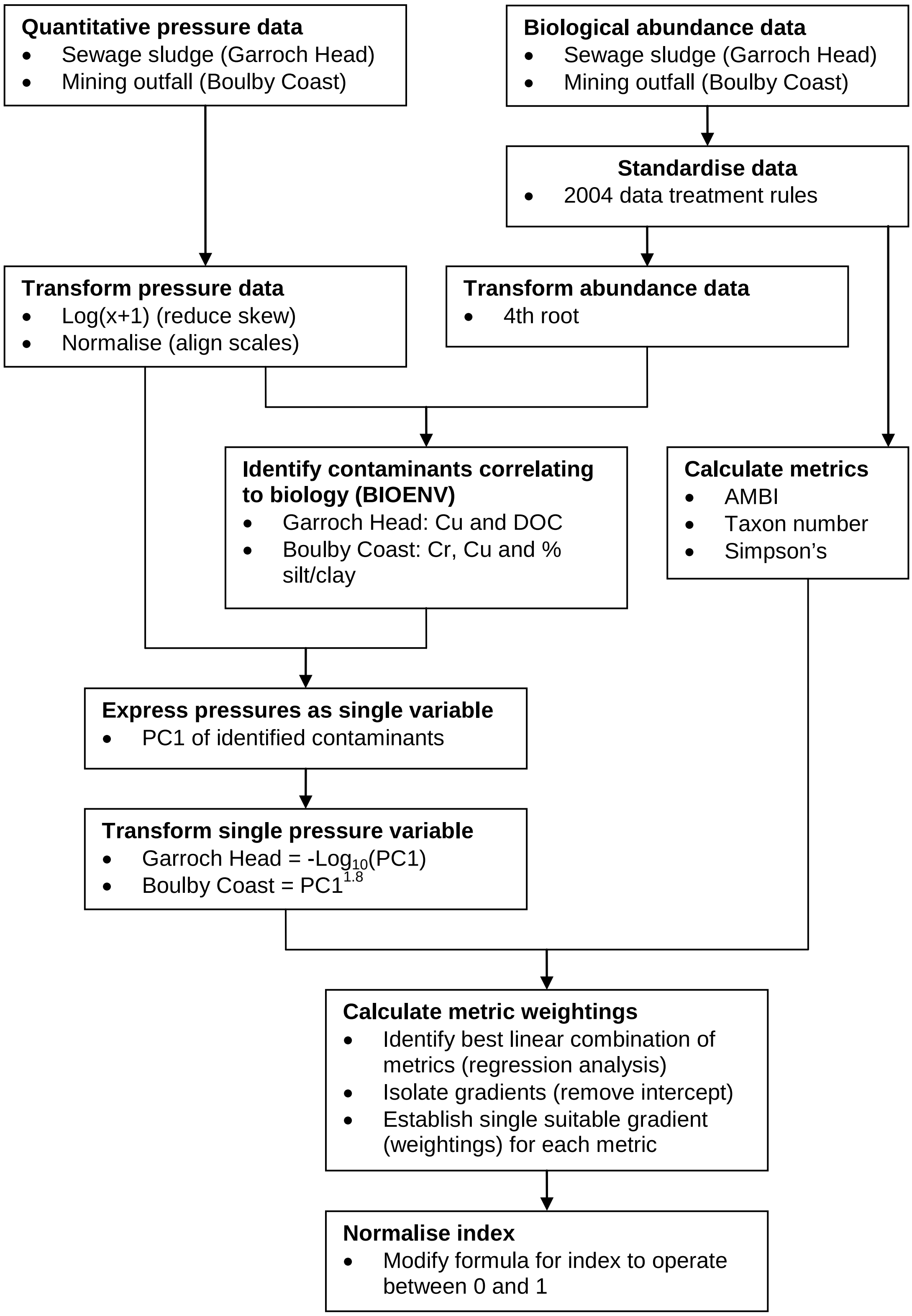


Figure 3.16 Process chart for the development of $\text{IQI}_{\text{v.IV}}$

Note: DOC, dissolved organic carbon.

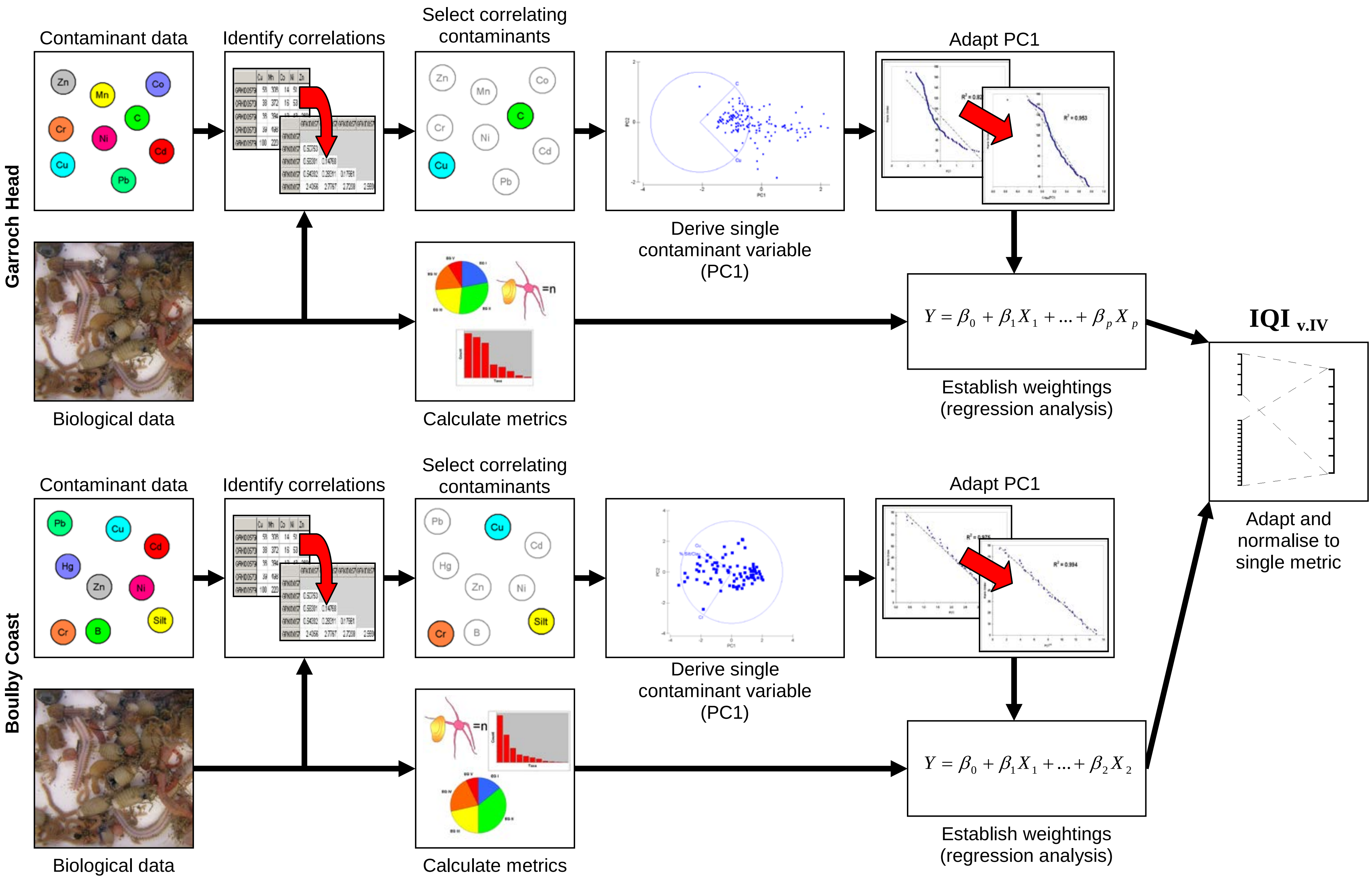


Figure 3.17 Illustration of the process undertaken in the development of $IQI_{v.IV}$

3.8.1 Refinement of the IQI metric weightings using regression analysis

A review of $IQI_{v.III}$ was undertaken in March 2006 with Bob Clarke, Richard Warwick and Paul Somerfield of Plymouth Marine Laboratories. As a result of this external consultation, the IQI was revised following the recommendation to use regression analysis to establish the appropriate weightings of the component metrics.

Regression analysis enables the identification of relationships between a response and one or more predictors. In this case, the objective was to use the measured biological indicators (that is, metrics) to predict or estimate the extent of anthropogenic pressure. The data used in the regression analysis were the biological metrics and quantitative pressure data (predictors and response respectively) from the Garroch Head and Boulby Coast pressure gradients (Section 3.3). Once the new weightings were established, the resultant index was tested against further datasets.

Regression analysis involves the identification of the relationship between a dependent variable and one or more independent variables. To revise the IQI, regression analysis was used to identify the weightings of the individual metrics that maximised the linear correlation between $IQI_{v.IV}$ and a function of pressure. The multiple linear regression model for the development of $IQI_{v.IV}$ is expressed as follows:[7]

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p \qquad \text{Equation 3.11}$$

where:

Y = dependent/response variable (pressure)

X = independent/predictor variable (metric)

β_0 = constant coefficient

β = predictor variable coefficient

Substituting the metrics into the above model resulted in the following formula:

$$f(\text{Pressure}) = \beta_0 + \beta_1 (AMBI_{IQI}) + \beta_2 (1 - \lambda'_{IQI}) + \beta_3 (S_{IQI})^* \qquad \text{Equation 3.12}$$

* Transformation providing highest correlation (see below)

In using regression analysis to develop a model to predict the extent of pressure from a selection of metrics, the biological metrics were set as independent/predictor variables with a function of the pressure set as the dependent/response variable.

The coefficients calculated from the regression analysis formed the weightings for the $IQI_{v.IV}$ metrics.

Further adaptation of the regression formula was then necessary to ensure the final IQI values operated over the required zero to one EQR scale.

Before the weightings could be established using regression analysis, it was necessary to adapt the expression of the metric S/S_{Ref}. $IQI_{v.III}$ incorporated a function of S/S_{Ref} as a multiplier. However, in regression analysis, the predictors (S/S_{Ref} and

[7] The regression analysis was based on the ordinary least squares method where the assumption of exogeneity is applied (residual error mean of zero).

other metrics) are multiplied by the coefficient values (rather than being multipliers themselves) so the inclusion of S/S_{Ref} as a multiplier was incompatible with the process. As described in Section 3.5, S/S_{Ref} was incorporated in $IQI_{v.III}$ in a form that down weighted the effect of variations in S/S_{Ref} observed at high values of S/S_{Ref} so that IQI values are only significantly affected when S/S_{Ref} reaches low values.

To satisfy the inclusion of S/S_{Ref} as an additive function (thus fulfilling the requirement of the regression analysis approach), transformation of the term was required. The most appropriate transformation of S/S_{Ref}, when included as a predictor in regression analysis, maximises the linear relationship between the metrics (predictors) and the pressure (response).

To identify an appropriate transformation of S/S_{Ref} for linear regression, the correlation between pressure and a range of power transformations of $(S/S_{Ref})^x$ was undertaken (x varied at 0.1 intervals between 0.01 and 1). The effect on the linear correlation between $(S/S_{Ref})^x$ and pressure gradient with changing values of x (power) can be observed in the Garroch Head data (Figure 3.18) and the Boulby Coast data (Figure 3.19).

For both datasets, the strong transformation of S/S_{Ref} served to increase the correlation between the pressure variable and the metric. Incorporation of taxa number in the form $S/S_{Ref}^{0.1}$ was adopted.

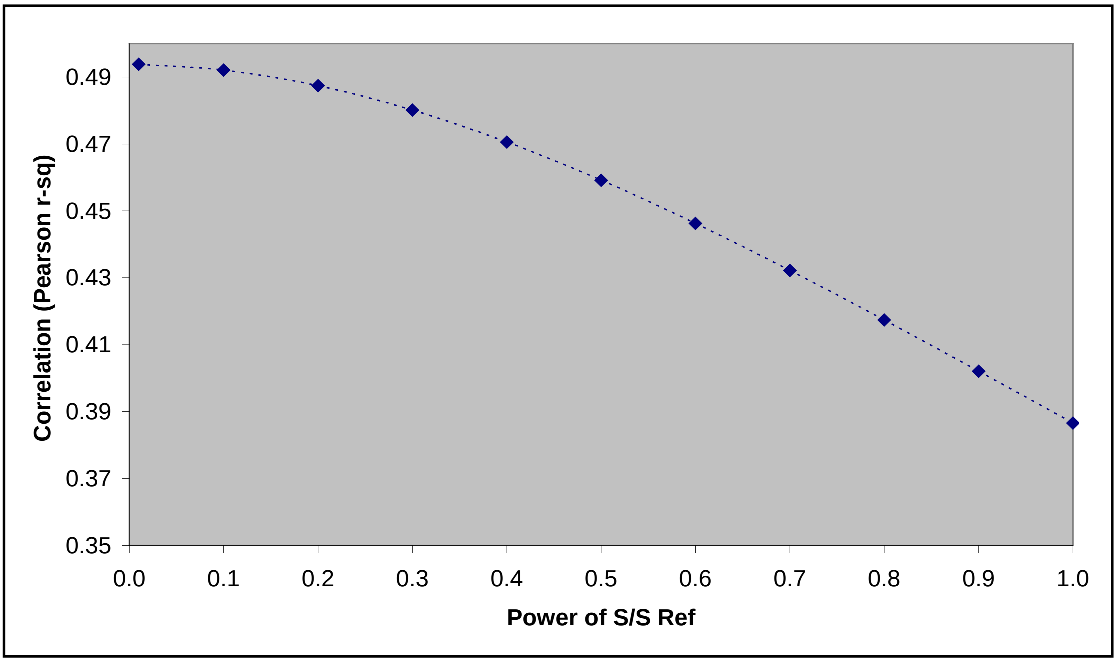


Figure 3.18 Correlation between different power transformations of S/S_{Ref} and pressure gradient (Section 3.8.2) in the Garroch Head Sewage Sludge disposal ground data

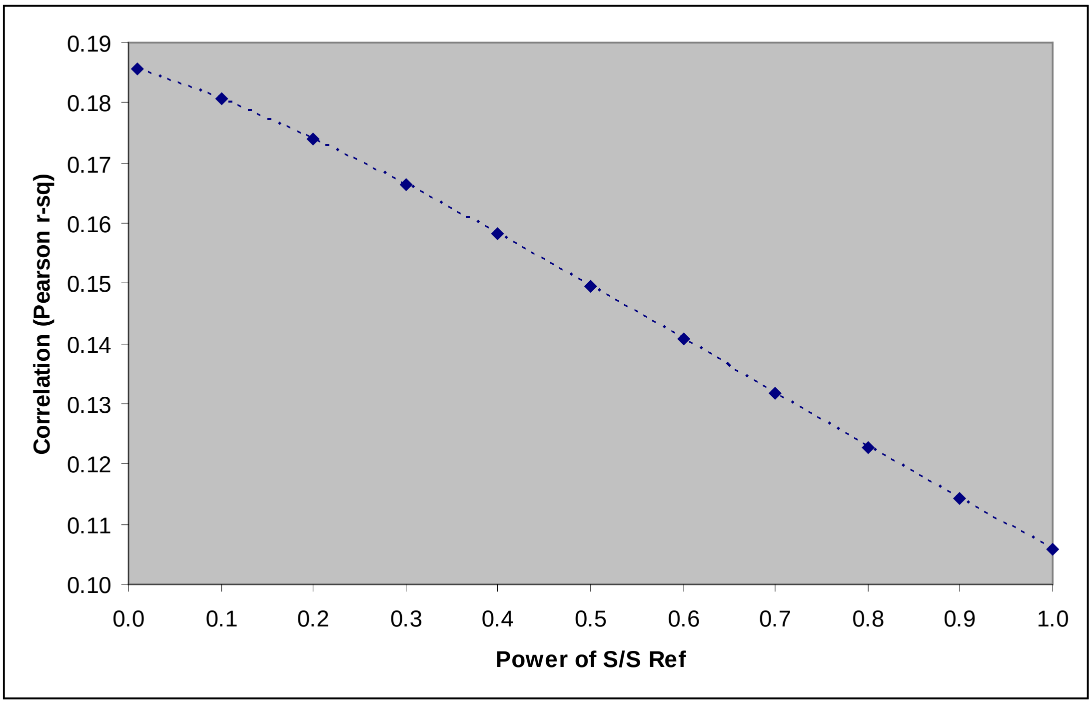


Figure 3.19 Correlation between different power transformations of S/S_{Ref} and pressure gradient (Section 3.8.2) in the Cleveland Potash mining waste disposal ground data

3.8.2 Representing pressure as a single variable

Regression analysis can only identify the coefficients for the predictors (metrics) for a single response (pressure). However, in both the Garroch Head and Boulby Coast studies, the variability of the biological assemblages correlated with the variability of multiple contaminants (pressures). As regression analysis requires pressure to be expressed as a single response variable, the variability of the multiple pressures needed expressing as a single variable. The approach adopted was based on a two-step analysis of the data as follows:

- **Step 1:** Identify those contaminants that correlated to the variability of the benthic assemblages over the pressure gradient (BIOENV)
- **Step 2:** Derive a single proxy variable that best explained the overall variability over the pressure gradient of those multiple contaminants identified through Step 1 (PCA)

Analysis of the data through BIOENV and PCA was applied as described below.

BIOENV analysis

The BIOENV routine within the PRIMER© version 6 statistical software package (Clarke and Warwick 2001) allows for the non-parametric comparison of multivariate biological and environmental datasets. BIOENV can be used to identify, from a range of variables, a single or subset of those variables that have the highest rank order correlation with the variability of a second set of variables across a dataset. In this instance, the routine was used to identify those measured contaminants that had the highest correlation with the variability of biology abundance data. For the analysis, contaminant data was log(x+1) transformed and then normalised. The instances of high abundance in the biology data (>1,000 for certain taxa) resulted in 4th root transformation being applied.

For Garroch Head, BIOENV analysis identified Cu and dissolved organic carbon (DOC) as the contaminants with the highest correlation to the variability of the benthic abundance data (Spearman rank correlation (ρ_w) of 0.645). For Boulby Coast, BIOENV analysis identified Cu, Cr and percentage silt/clay as the contaminants with the highest correlation to the variability of the benthic abundance data (ρ_w of 0.616). The contaminants identified through BIOENV were then taken forward for PCA.

Principal component analysis

PCA is a process whereby sample data for multiple variables (high-dimensional variable space) are projected onto a low number of specified axes, or principal components (PCs). Each principal component provides in effect an axis of best fit through the multiple variables; the first principal component (PC1) being the axis that explains the most variability throughout the data, with subsequent axes explaining diminishing levels of variability. PC1 provided a suitable proxy variable that expressed the overall variability of the pressure gradient data as a single variable representative of those contaminants that in combination have the highest correlation with the benthic abundance data as identified through BIOENV.

Contaminant data was log(x+1) was transformed and then normalised for the PCA. The PC1 of the Garroch Head data explained 91.1% of the variation of the transformed Cu and DOC data over the pressure gradient. For the Boulby Coast data, the PC1 data explained 80.2% of the variation of the transformed Cu, Cr and percentage silt/clay data over the pressure gradient.

PC1 adaptations

The approach aimed to identify a single model (linear combination) of taxa number, 1-(AMBI/7) and 1-λ′ that was a compromise of the two models that best fit the gradient of the Garroch Head and Boulby Coast pressure data. The use of the two separate pressure gradient datasets in the formulation of the IQI therefore required the pressure data to be transformed to increase the comparability of the gradients.

As the contaminant data of the two datasets (and corresponding first principal components) do not necessarily operate over comparable scales (that is, the rate of increase of the contaminants over the gradient may differ), this may have increased the differences between the linear relationships of the metric/pressures (for example, one dataset may have had a more curvilinear relationship than the other) unless they are transformed. In terms of using a function of the pressure data as the response variable to be 'predicted' by taxa number, 1-(AMBI/7) and 1-λ′ through multiple linear regression, using the raw PC1 has the advantage of preserving the relative distance between samples in terms of the contaminant data. However, the rate of change in the PC1 from the least to most contaminated samples (that is, the curvilinearity of the PC1 over the gradient) cannot be assumed to be the same for the two datasets.

The use of PC1 rank scores as the response variable does not preserve the relative distance between samples in the analysis in terms of the pressure data. Their use does, however, have the advantage when comparing the two datasets in that the rate of change in the rank scores from the least to most contaminated samples will be the same for the two datasets.

To derive a function of the pressure data that both preserved the relative distance between samples in terms of levels of contaminants (PC1) yet attempted to increase the comparability of the two pressure data gradients, the pressure data PC1 values were transformed to increase their linear correlation with the rank order of the PC1 in order to align the pressure scales for the regression analysis. Various transformations of the PC1 for each dataset were trialled with the highest correlation between PC1 and rank order PC1 being attained with $-\log_{10}$(PC1) and $PC1^{1.8}$ for the Garroch Head (Figure 3.20) and Boulby Coast (Figure 3.21) datasets respectively.

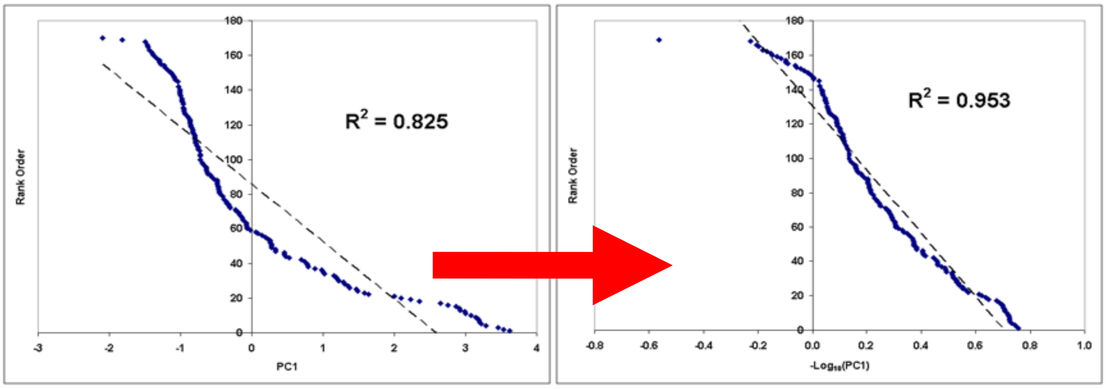


Figure 3.20 Effect of transformation (-log10) of PC1 sample values for the Garroch Head disposal ground contaminant data in terms of correlation to the rank order (increase of R^2 from 0.825 to 0.953)

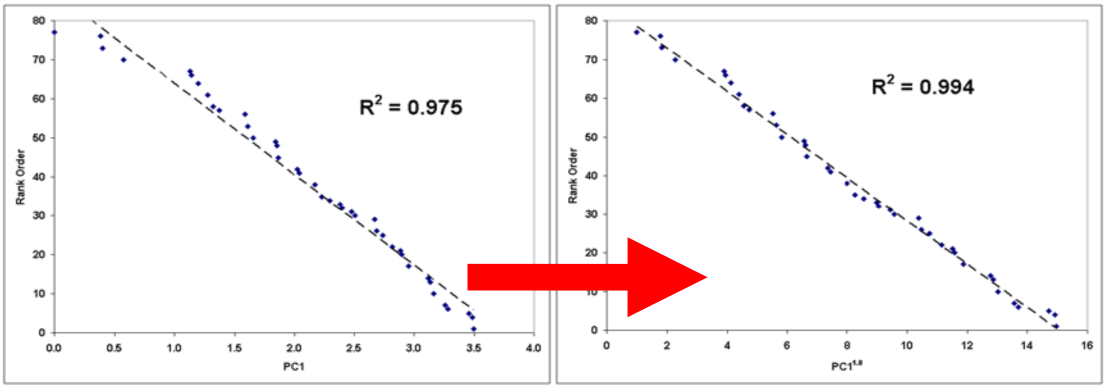


Figure 3.21 Effect of transformation (PCA1.8) of PC1 sample values for the Boulby Coast contaminant data in terms of correlation to the rank order (increase of R_2 from 0.975 to 0.994)

3.8.3 Regression analysis results

Regression analysis was performed using Minitab© statistical software. As the objective was to use the selected biological metrics to estimate the effect of pressure, the biological metrics were used as predictor variables with the function of pressure data PC1 being included as the response variable. The results of the analysis for the two pressure gradient datasets were as described below.

Garroch Head

Applying the coefficients for each metric (Table 3.2), enables the regression formula for the Garroch Head dataset to be derived:

$$-Log_{10}(PC1) = -1.43 + (0.528 \times AMBI_{IQI}) + (0.168 \times 1 - \lambda'_{IQI}) + (0.801 \times S_{IQI}^{\ 0.1})$$

Equation 3.13

Table 3.2 Results from regression analysis between Garroch Head pressure data (–log10(PC1)) and IQI metrics

Predictor	Coefficient	SE coefficient	T	Probability
Constant	-1.425	0.152	-9.37	<0.001
$AMBI_{IQI}$	0.528	0.052	10.10	<0.001
$1-\lambda'_{IQI}$	0.168	0.071	2.36	0.019
$S_{IQI}^{\ 0.1}$	0.801	0.206	3.90	<0.001
S = 0.140				
R^2 = 71.4%				

Notes: SE = slope standard error coefficient
T = Student's t-test statistic (test against threshold to determine significance)
S = square root of mean square error
R^2 = percentage of response variable (pressure) variation explained by its relationship with the predictor variables (metrics).

The regression explained 71.4% of the error within the Garroch Head pressure data (Table 3.3).

Table 3.3 Analysis of variance results between Garroch Head pressure data (–log10(PC1)) and regression

Source	DF	SS	MS	F	Probability
Regression	3	8.340	2.780	142.18	<0.001
Residual error	171	3.343	0.020		
Total	174	11.683			

Notes: DF = degrees of freedom
SS = sum of squares (error) between groups (source = regression) and within groups (source = residual error),
MS = mean square (SS/DF)
F = test statistic for comparison against threshold to determine significance

Boulby Coast

Applying the coefficients for each metric (Table 3.4) enables the regression formula for the Boulby Coast dataset to be derived:

$$PC1^{1.8} = -10.7 + (9.80 \times AMBI_{IQI}) + (-0.04 \times 1 - \lambda'_{IQI}) + (12.9 \times S_{IQI}^{\ 0.1})$$

Equation 3.14

Table 3.4 Results from regression analysis between Boulby Coast pressure data (PC11.8) and IQI metrics.

Predictor	Coefficient	SE Coefficient	T	Probability
Constant	-10.726	4.323	-2.48	0.015
$AMBI_{IQI}$	9.799	3.695	2.65	0.010
$1\text{-}\lambda'_{IQI}$	-0.043	1.611	-0.03	0.979*
S_{IQI}^$^{0.1}$	12.909	6.449	2.00	0.049
S = 3.380				
R^2 = 34.1%				

Notes: *Not significant (*p* >0.05)
SE = slope standard error coefficient
T = Student's t-test statistic (test against threshold to determine significance)
S = square root of mean square error
R^2 = percentage of response variable (pressure) variation explained by its relationship with the predictor variables (metrics).

The regression explained 34.1% of the error within the Boulby Coast pressure data (Table 3.5).

Table 3.5 Analysis of variance results between Boulby Coast pressure data (PC11.8) and regression

Source	DF	SS	MS	F	Probability
Regression	3	443.050	147.680	12.930	<0.001
Residual error	75	856.690	11.420		
Total	78	1299.740			

Notes: DF = degrees of freedom
SS = sum of squares (error) between groups (source = regression) and within groups (source = residual error),
MS = mean square (SS/DF)
F = test statistic for comparison against threshold to determine significance

Simpson's evenness (1-λ′) was shown to have no significant contribution in the multiple regression analysis of the Boulby Coast data at a 95% confidence level (T = -0.030 with associated probability of 0.979). The metric was retained for inclusion in the IQI due to its significant contribution in the Garroch Head regression analysis, although the ability of the metric to explain a low proportion of the error within the pressure data in contrast to S and 1-(AMBI/7) is reflected in its overall weighting within $IQI_{v.IV}$ (Section 3.8.4).

3.8.4 Adapting regression equations for the IQI

The regression intercepts were removed to isolate the gradients between the metrics and pressure data. As each metric is expressed as a ratio of observed to expected, with a minimum and maximum of zero and one respectively, the intercept can be considered as a redundant component of the equation.

To ensure that the maximum EQR derived from the IQI is equal to one (that is, when each observed metric value is equal to its value under reference conditions), the weightings needed to be normalised so that the sum of the coefficients was equal to one while retaining the relative weighting of each metric within each regression formula. This was achieved for each regression formula by dividing each metric coefficient by the sum of the coefficients for all metrics. Applying these rules to the Garroch Head and Boulby Coast regression formulae provided the normalised weightings as follows:

$$Garroch\,Head = \left(0.353 \times AMBI_{IQI}\right) + \left(0.112 \times 1 - \lambda'_{IQI}\right) + \left(0.535 \times S_{IQI}{}^{0.1}\right)$$

Equation 3.15

$$Boulby\,Coast = \left(0.432 \times AMBI_{IQI}\right) + \left(-\,0.002 \times 1 - \lambda'_{IQI}\right) + \left(0.569 \times S_{IQI}{}^{0.1}\right)$$

Equation 3.16

The approach was based on the assumption that the true range of pressure experienced by the benthic assemblages of both the Garroch Head and Boulby Coast are comparable, with comparably low levels of pressure at the reference condition sites. However, it is acknowledged that it is not known whether the degree of pressure at these sites is truly comparable (reference conditions are discussed in Chapter 4).

The objective was to develop a single classification tool to reflect overall anthropogenic disturbance in multiple pressure systems. The tool therefore needed to incorporate a single set of weightings for each metric that was a balance of those most suitable for correlating to pressure from the multiple pressure datasets available for analysis.

A weighted average of the gradients from the Garroch Head and Boulby Coast regression analysis were used to derive the weighting of each metric within the IQI. Rather than basing the weighting on the mean weight of each metric from the two datasets (for example, weight for AMBI = (0.353+0.432)/2), a weighted average was used. This was because the metric weightings from the Garroch Head analysis being identified as more effective predictors of a pressure gradient. As the metrics explained a lower degree of total error within the Boulby Coast pressure data than within the Garroch Head data, it was decided that the influence of the Boulby Coast weightings should be lower than the Garroch Head weightings. This approach operates on the principle that, if two hypothetical regression gradients provided R^2 values of for example, 90% and 10%, it would be more appropriate to weight the IQI metrics in accordance with the model that explained the greatest proportion of the total error in the pressure data.

The metrics were weighted according to the total error explained (R^2) by the corresponding regression formula as follows:

$$metric\;\;weighting = \frac{\left(\left(W_{GH} \times R_{GH}^2\right) + \left(W_{BC} \times R_{BC}^2\right)\right)}{R_{GH}^2 + R_{BC}^2}$$

Equation 3.17

where:

W_{GH} = normalised metric weighting from Garroch Head regression

W_{BC} = normalised metric weighting from Boulby Coast regression

R^2_{GH} = error (R^2) from Garroch Head regression

R^2_{BC} = error (R^2) from Boulby Coast regression

Applying the approach to the IQI metric values resulted in the following weightings:

$$AMBI_{IQI}\,weighting = \frac{((0.353 \times 71.4) + (0.432 \times 34.1))}{71.4 + 34.1} = 0.38$$

$$1 - \lambda'_{IQI}\ weighting = \frac{((0.112 \times 71.4) + (-0.002 \times 34.1))}{71.4 + 34.1} = 0.08$$

$$S_{IQI}{}^{0.1} weighting = \frac{((0.535 \times 71.4) + (0.569 \times 34.1))}{71.4 + 34.1} = 0.54$$

Applying these weightings resulted in the following single regression equation:

$$Multimetric = (0.38 \times AMBI_{IQI}) + (0.08 \times 1 - \lambda'_{IQI}) + (0.54 \times S_{IQI}{}^{0.1}) \qquad \text{Equation 3.18}$$

However, the minimum theoretical metric values obtained from the multimetric in Equation 3.18 are not zero. To modify the equation so that the IQI operates between zero and one, normalisation was required. This was undertaken as follows:

$$IQI = \frac{Int_{Obs} - Int_{Min}}{Int_{Max} - Int_{Min}} \qquad \text{Equation 3.19}$$

where Int = intermediate multimetric value.

As azoic samples need to be treated separately (see Section 3.9), minimum metric values needed to be established as the lowest values possible where samples are not azoic. For $AMBI_{IQI}$, the minimum value is where all taxa are in ecological group V (first order opportunistic species). This equates to a minimum $AMBI_{IQI}$ value of 0.143. The minimum taxa number is 1. Values of S (following data standardisation) in the WFD R&D database (2006) had an upper limit of approximately 100 taxa per 0.1 m^2. Where S_{Ref} equals 100 and S equals 1, the value of $S_{IQI}{}^{0.1}$ is 0.631.

To identify the minimum theoretical value of the intermediate formula (Int_{min} in Equation 3.19), the weightings from the regression were applied to the minimum values for $AMBI_{IQI}$ and $S_{IQI}{}^{0.1}$ (Equation 3.20). The minimum value for Simpson's evenness (expressed as 1-λ′) is 0 (all individuals are the same taxa) and as such does not need to be factored into the scaling process.

$$Int_{Min} = (0.38 \times (1 - (AMBI_{Max}/7))) + (0.54 \times (S_{Min}/S_{Ref})^{0.1}) \qquad \text{Equation 3.20}$$

where:

$AMBI_{Max}$ = maximum of AMBI (results in minimum 1-(AMBI/7))

S_{Min} = minimum value of S in non-azoic sample

By substituting the values of $AMBI_{Max}$ and S_{Min} in Equation 3.20, Int_{Min} was calculated as:

$$(0.38 \times 0.143) + (0.54 \times 0.631) = 0.40$$

Equation 3.21

The maximum value for the intermediate multimetric is where all metrics are equal to their corresponding reference condition values and so the maximum multimetric value is set to one. To normalise the index, minimum and maximum multimetric values were applied as follows:

$$IQI = \frac{Int_{Obs} - 0.4}{1 - 0.4}$$

Equation 3.22

By applying the metric weightings and normalising to operate fully between zero and one as described above, the $IQI_{v.IV}$ is expressed as follows:

$$IQI_{v.IV} = \left(\left(0.38 \times \left(\left(AMBI_{IQI}\right)\right)\right) + \left(0.08 \times \left(1 - \lambda'_{IQI}\right)\right) + \left(0.54 \times \left(S_{IQI}\right)^{0.1}\right) - 0.4\right) / 0.6$$

Equation 3.23

3.9 Additional factors for consideration

Several other factors required consideration separately from to the classification tool development process described above.

3.9.1 Treatment of azoic samples

Azoic[8] samples are assigned EQR values of zero. The true significance of an azoic sample in terms of assessing the ecological health of a benthic assemblage will, in reality, depend on the natural conditions of the sample. An azoic sample from a typically species rich habitat (for example, mixed subtidal marine substrata) may be considered more indicative of anthropogenic pressure than an azoic sample from a naturally taxon-poor habitat (for example, sandy habitats in high energy variable salinity environments). If an azoic sample is assigned an EQR of greater than zero this limits the lowest ecological status to which a sample can be assigned; the classification tool would not be able to distinguish between, for example, poor and bad status (the lowest attainable status would be poor).

3.9.2 Unassigned AMBI ecological groups

The AMBI depends on the assignment of sensitivity scores to the taxa within a benthic assemblage in accordance with their assigned ecological group. The AMBI protocol recommends that where >20% of taxa are not assigned an ecological group, samples should be evaluated with care and should be excluded from analysis where >50% are unassigned (Borja and Muxika 2005). High proportions of unassigned taxa reduce the reliability of the assemblage AMBI value, potentially resulting in greater variability in AMBI. However, while potentially more variable, samples with a high proportion of unassigned taxa still provide additional information to assist with, for example, the identification of relationships between AMBI values and anthropogenic

[8] For the purpose of the WFD benthic classification, the term 'azoic' is used relative to the macrobenthic invertebrate fauna retained after processing and following taxon standardisation. It is acknowledged that samples described as azoic may in reality contain fauna not classified as macrobenthic invertebrates (for example, meiofauna, fish).

pressure gradient data. Samples with high proportions of taxa with unassigned AMBI ecological groups are also of value in relating anthropogenic pressure to the metrics S and 1-λ′. Therefore, for the purpose of classification tool development, all samples (including those with >50% of taxa unassigned to an AMBI ecological group were retained in the analysis.

3.9.3 Influence of habitats

The physicochemical conditions underlying the pressure gradient datasets are from a limited spectrum of habitats relative to the full range of habitats to which IQI assessments will be trialled and potentially used. The correlation between the metrics as weighted within $IQI_{v.IV}$ and comparable pressures within a broader range of habitats (substrata and salinity regimes) may not necessarily be the maximum possible. Metric selection and weighting would need to be established using pressure gradient data from a broader range of habitats to ensure that the weightings are optimal for all scenarios. It is recognised that the weightings derived from the test data are likely to be suboptimal for habitats outside those within the Garroch Head (fully marine subtidal mud) and Boulby Coast (fully marine subtidal fine sand) datasets. The influence of these weightings will continue to be evaluated as a broader range of habitats are incorporated into the assessment process.

3.9.4 Additional pressures

The current metric selection and weighting within $IQI_{v.IV}$ is based on data from a limited range of types of pressure (organic enrichment, hazardous substances and physical smothering). As with the range of habitats, the range of pressures experienced by macrobenthic assemblages in the marine environment extends beyond that in the test data (for example, dredging, navigation disturbance and trawling). Equally, water bodies (particularly transitional waters) are often subjected to multiple pressures. The selection of metrics, with corresponding weighting and transformation within the IQI, is also likely to not have the maximum correlation to pressures in all cases. Further testing will be carried out to establish the effectiveness of the IQI within a broader range of habitats and to a wider range of pressures (including multiple pressures).

The response of the AMBI to different anthropogenic pressures is a crucial factor in the response of the IQI to such pressures. The AMBI was initially developed by Borja et al. (2000) using the response of taxa with different sensitivities over an organic enrichment pressure gradient. However, the response of AMBI to a broad range of pressures has since been tested in numerous studies, having been demonstrated to respond to pressures such as sediment extraction, engineering works, metal contamination, oil extraction activity and anoxia. The applicability of AMBI to different impact sources is illustrated by Muxika et al. (2005).

3.9.5 Exceeding reference conditions

As reference conditions are not based on absolute maximum metric values in the available data (see Chapter 4), there is potential for observed metric values to exceed those expected under reference conditions (observed/expected >1). Therefore there is the potential for the corresponding EQR to exceed one. At an EQR of zero, all values will be the lowest value achievable for a given metric (on the basis of the sample not being azoic).

4 Reference conditions

Reference conditions provide the benchmark against which the ecological health of observed samples is assessed under the WFD. This chapter describes how reference conditions were established for the IQI.

4.1 Incorporation of reference conditions in the IQI

The establishment of type-specific reference conditions is central to the status assessment of biological quality elements in the WFD:

> 'Type specific biological reference conditions shall be established, representing the values of the biological quality elements … for that surface water body type at high ecological status' (WFD, Annex II section 1.3(i)).

The WFD normative definitions provide a qualitative description of the benthic invertebrate assemblage at high status, the upper end of high status equating to reference conditions:

> 'The level of diversity and abundance of invertebrate taxa is within the range normally associated with undisturbed conditions. All disturbance-sensitive taxa associated with undisturbed conditions are present' (WFD, Annex V section 1.2).

Observed assemblages are compared with the benchmark 'reference condition' descriptions to give an assessment of ecological status. As such, the EQR describes the numerical relationship between the observed value and the reference condition value. The EQR scale ranges from zero to one, with reference conditions represented by an EQR of one.

$IQI_{v.IV}$ describes each metric as a weighted proportion of the ratio of observed/expected values (see Chapter 3):

$$IQI_{v.IV} = \left(\left(0.38 \times \left(\frac{1-(AMBI/7)}{1-(AMBI_{\mathrm{Ref}}/7)} \right) \right) + \left(0.08 \times \left(\frac{1-Lambda'}{1-Lambda'_{\mathrm{Ref}}} \right) \right) + \left(0.54 \times \left(\frac{S}{S_{\mathrm{Ref}}} \right)^{0.1} \right) - 0.4 \right) \Big/ 0.6$$

Equation 4.1

where:

Ref = expected metric value under reference conditions.

As such, reference conditions are incorporated at the metric level. Where observed values are equal to expected values, the EQR is equal to one.

The extent of departure of an observed metric from its expected reference value is directly proportional regardless of the habitat and methods that the observed sample originates from. For example, the ecological status for a sample where the observed taxa number is 25 and expected taxa at reference condition is 75 is equivalent to the ecological status where the observed taxa number is 15 and expected taxa at reference condition is 45 (Figure 4.1).

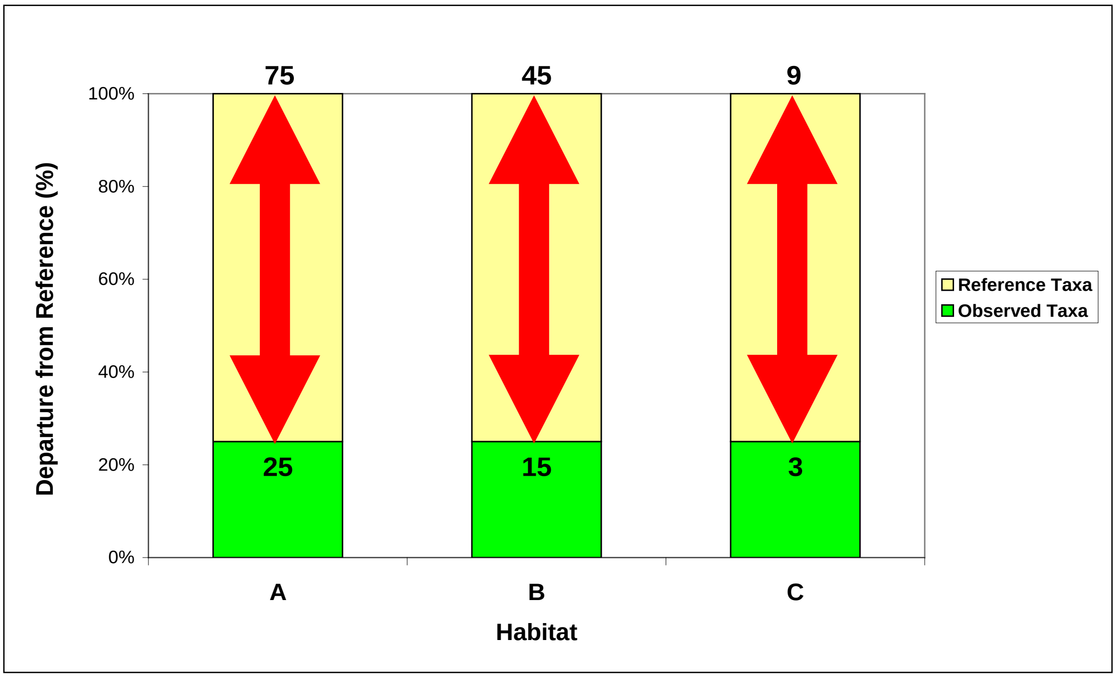


Figure 4.1 Values of observed taxa from different habitats, representing comparable relative departure of observed taxa from reference conditions

The establishment of appropriate reference conditions is therefore critical for the assessment of all water bodies and quality elements. Although reference conditions are not necessarily the final targets for all water body components, they provide the direction for improving ecological restoration (COAST 2003).

4.2 Options for setting reference conditions

The WFD states that reference conditions may be based on either spatially derived conditions or modelling, or a combination of both (WFD, Annex II, section 1.3(iii)). Where these methods cannot be used, expert judgement may be applied to establish reference conditions. The WFD CIS Guidance Document No. 5 (COAST 2003) lists the method options for setting reference conditions in the following hierarchical order:

1. Data from an existing undisturbed site or a site with only very minor disturbance
2. Historical data and information
3. Models
4. Expert judgement

The advantages and disadvantages of each option were considered in order to develop the most appropriate approach to setting IQI reference conditions (Table 4.1). It was decided to set reference conditions following the approach whereby a combination of expert judgement and data from existing undisturbed sites or sites with minor disturbance was used to initially set values for a single habitat and set of sample collection and processing methods.

Models were developed to adapt the initial set of reference condition values for a broader range of different habitats and methods.

Table 4.1 Advantages and disadvantages of the four options for setting reference conditions as described by WFD CIS Guidance Document No. 5 (COAST 2003)

Option	Advantage	Disadvantage
Existing undisturbed site	• Potential to use approach to identify true non-impacted reference conditions • Reference conditions can be used as found in data – no need to interpret or adapt to accommodate potential between habitat or method differences in values	• Subjectivity in identifying sites in absence of quantitative pressure data • Low data availability from true undisturbed conditions for the required habitats and sampling methods • Reference conditions may be absent for certain habitats and sample methods
Historical data and information	• Potential for the approach to provide information on true non-impacted conditions • Large quantity of historical benthic macroinvertebrate data • Broad range of habitats and sampling methods in existing data	• Subjectivity in identifying sites in absence of quantitative pressure data • Reliant on existence of suitable data for all habitats and methods currently used for classification • Reference conditions may be absent for certain habitats and sample methods • Biological data often poorly supported by environmental data • Scope for inconsistency between reference conditions from different habitats and sample methods • Historic data subject to elevated degrees of inconsistency (for example, QA and AQC procedures, taxonomy, habitat definitions, and so on)

Models	• Setting of values repeatable • Between-habitat differences can be established to be proportional to trends in real data • Allows the extrapolation of reference conditions for habitats from existing data where gaps exist • Potential to ensure reference conditions change proportionally to methodology induced bias	• Models are limited to the extent to which true non-impacted data exists in the available data upon which the models are based – approach may lead to underestimations of reference conditions • Large datasets required to attain high levels of confidence in reference condition values (subject to variability in biological and environmental data) • Effectiveness of models limited by the underlying empirical evidence
Expert judgement	• Potential for informed decisions to establish true non-impacted reference conditions • Understanding of the extent to which metrics are influenced by habitats may enable reference conditions to be set for the full range • Low data requirements – possible to apply qualitative/anecdotal information • Understanding of the influence of changing methods to metrics • Data from existing programmes sufficient to support informed expert judgement	• Extent to which reference conditions are influenced by habitat and sampling collection and processing methods is open to subjectivity – scope for inconsistency between reference conditions for different habitats and methods

4.3 Factors influencing metric reference condition values

The metrics used to describe the structure and function of macrobenthic assemblages are influenced by a multitude of factors that are independent of anthropogenic disturbance. These factors may originate from:

- true differences in the data (that is, changes to the assemblages due to differing environmental conditions such as salinity)

- artefacts in the data (that is, changes as a result of how the assemblages are observed such as sample collection and processing methods).

Corresponding reference condition values are expected to be similarly influenced. The metric reference condition values for the IQI therefore need to be adapted to ensure that the influences of habitat and sample collection and processing method are not misinterpreted as anthropogenic disturbance.

UK transitional and coastal waters are made up of a broad range of habitats, including gradients of salinity (~0.5 to ~34) and various sediment types (silt/clay, gravels, mixed sediments and so on). Individual water bodies generally consist of a mosaic of different habitats (in terms of substrata, intertidal/subtidal areas and salinity), which in turn may require different sampling collection and processing methods. For any sample, it is crucial that the physicochemical conditions and sampling methods are known in order to compare the observed sample with the appropriate expected reference values when calculating the EQR. Without such supporting information, the observed metrics may be interpreted out of context, resulting in misinterpretation of the degree of anthropogenic disturbance experienced by the macrobenthic assemblage.

Unless it can be demonstrated that a biological metric is either independent of physicochemical parameters and sampling methodologies, or responds differently depending on whether pressures are anthropogenic or natural, such supporting information must be factored into the assessment to ensure a minimal response of the metric attributable to non-anthropogenic pressures.

4.3.1 Effect of natural environmental conditions

Factors such as salinity regime and aerial exposure act as natural pressures on macrobenthos, with a potential consequence being the absence of taxa that are directly intolerant of, or poor competitors when exposed to, such conditions. The behaviour of the IQI metrics for assemblages exposed to such conditions tends to be comparable with assemblages exposed to anthropogenic pressure, that is, a reduction in taxa numbers and increasing dominance of opportunistic taxa (reducing values of 1-AMBI/7 and 1-λ′).

According to Elliott and Quintino (2007), the extent of influence of natural pressures is particularly important in transitional waters where:

> '… the dominant estuarine faunal and floral community is adapted to and reflects high spatial variability in naturally highly stressed areas but that it [the community] has features very similar to those found in anthropogenically-stressed areas thus making it difficult to detect anthropogenically-induced stress in estuaries'.

However, the underlying principle of the Estuarine Quality Paradox may also be relevant to other aspects of the physicochemical environment. Sediment characteristics affect the percolation properties of a substratum, so substrata dominated by silt/clay in low energy environments may experience poor water and dissolved oxygen circulation, resulting in naturally anaerobic conditions. The natural faunal assemblage associated with such sediments may have elevated numbers of r-strategist taxa tolerant to such anaerobic conditions, resulting in AMBI values appearing more representative of anthropogenically disturbed conditions. Conversely, mixed substrata hold a greater variety of niches (increased habitat complexity) and therefore naturally elevated numbers of taxa in contrast to homogenous substrata.

To mitigate the impact of these natural environmental conditions, the IQI utilises reference conditions that are set according to the specific conditions with which the biological assemblage is associated (see Section 4.6).

4.3.2 Effect of sample collection and processing methods

The metrics within the IQI are influenced by different sample method attributes such as sample surface area (for example, increasing taxa number as area increases) and sample depth (for example, possible influences to AMBI from increased abundance of larger burrowing fauna such as pollution sensitive k-strategists with deeper core samples). For these reasons, reference condition values must be specific to the sample collection method. In certain cases, the close similarity between sampling methods (for example, 0.1 m^2 Day grab and 0.1 m^2 Van Veen) and the biological metrics derived from their use (see Proudfoot et al. 1997) may enable reference condition values to be applicable to different methods within a similar group. However, it is recommended that testing of the differences in the context of the IQI is undertaken before confirming the applicability of reference conditions to different sampling methods.

Marine benthic infauna are predominantly sampled using grabs or cores, generally targeting subtidal and intertidal habitats respectively. Consequentially, there may also be additional indirect effects of differing sample methods whereby certain physicochemical conditions associated with intertidal habitats (for example, aerial exposure and greater fluctuations in temperature) serve to increase levels of natural pressure, with reduced taxa numbers and elevated proportions of stress-tolerant taxa (influential to AMBI values) being observed in core data as a consequence.

As well as the collection method used, sample processing (for example, sieving, subsampling) can influence the retention of macrofauna (Lewis and Stoner 1981) and therefore corresponding metric scores. The effect of sieve mesh aperture on the IQI metrics was analysed using benthic infauna data collected for the UK's CSEMP. CSEMP benthic invertebrate samples, collected between 1999 and 2006, were each processed through 0.5 mm and 1 mm sieve meshes. As all other variables (methodological and physicochemical) were constant, the influence of sieve mesh aperture and its effect on each metric within the IQI could be quantitatively described. The majority of the data relate to 0.1 m^2 Day grab samples (subtidal), but a limited set of 0.01 m^2 hand core (intertidal) samples allowed both methods to be evaluated.

Prior to analysis, the IQI metric values were transformed to ensure normality of the data. Differences in IQI metric values (transformed) between 0.5 mm and 1 mm sieve mesh sizes were indicated using ANOVA (Table 4.2).

Table 4.2 ANOVA results for the comparison of IQI metrics (transformed) for benthic infaunal samples analysed through 0.5 mm and 1 mm sieve mesh apertures (grab and core samples)

Method	*n*	Metric	Mean value (0.5 mm)	Mean value (1 mm)	Probability
Grab (0.1 m^2)	866	Log(taxa number+1)	3.001	2.704	<0.001
		Arcsine (1-(AMBI/7))	0.536	0.578	0.001
		Arcsine (1-λ′)	0.783	0.750	0.022
Core (0.01 m^2)	135	Log(taxa number+1)	2.561	2.173	<0.001
		Arcsine (1-(AMBI/7))	0.328	0.397	0.002
		Arcsine (1- λ′)	0.695	0.705	0.723*

*Not significant (p >0.05).

For both grab and core methods the sieve mesh aperture corresponded to a significant (p <0.05) difference in taxa number, with greater numbers being retained on the 0.5 mm sieve mesh.

- The non-transformed mean taxa number for grabs was 19 and 14 for 0.5 mm and 1 mm mesh sizes respectively.
- The non-transformed mean taxa number for cores was 12 and 8 for 0.5 mm and 1 mm mesh sizes respectively.

Mean (1-(AMBI/7)) values differed significantly (p <0.05) between the mesh size fractions for data from both grab and core methods, indicating that the invertebrate composition of the >0.5 mm fraction had a lower proportion of sensitive taxa in contrast to the corresponding >1 mm fraction. This is consistent with the supposition that the taxa between 0.5 and 1 mm are generally opportunist taxa (r-strategists) in contrast to those >1 mm.

The effect of sieve mesh aperture on 1-λ′ for grab samples displays elevated evenness for >0.5 mm fractions in contrast to the >1 mm fraction. The core data indicates no significant difference in average evenness between the >0.5 and 1 mm fractions (p >0.05).

The differences in metric values observed between the >0.5 mm and >1 mm fractions need to be reflected in the expected values under reference conditions, that is, IQI metric reference values must be specific to the sieve mesh aperture.

4.3.3 Influence of data standardisation rules

Changes in data standardisation procedures (for example, removal of certain taxonomic groups) will have implications on the reported structure of a benthic assemblage and thus its associated metric values. As with the development of the IQI, data treatment rules have been developed in phases (see Chapter 2). These revisions had implications for the metric scores derived from abundance data and in turn required adaptation of the expected reference values.

4.4 Initial approach to setting $IQI_{v.IV}$ reference condition values (2006)

In the development of $IQI_{v.IV}$, the, reference conditions for coastal sublittoral sand and mud were based on the maximum values of S, 1-(AMBI/7) and 1-λ′ found within the Garroch Head sewage sludge disposal ground dataset (following standardisation) (Table 4.3).

Table 4.3 Maximum values for $IQI_{v.IV}$ metrics used as preliminary reference condition values for coastal sublittoral sand and mud (0.1 m^2 grab, 1 mm sieve mesh)

$IQI_{v.IV}$ metric	Reference condition value for fully marine subtidal fine sand/mud habitats (0.1 m^2, 1 mm sieve mesh)
Taxa number	82
1-(AMBI/7)	1.00
1-λ′	1.00

However, these reference condition values for the IQI metrics were revised by the UK and RoI WFD competent authorities in September 2006 using a combination of standardised data from the national monitoring programmes and expert judgement. Metric reference condition values were established using data conforming to the following criteria:

- minimal anthropogenic disturbance (expert judgement)
- coastal waters (salinity >30)
- subtidal
- fine sand/mud habitats (EUNIS type A5.2 and A5.3 habitats)
- 0.1 m^2 sample surface area
- 1 mm sieve mesh aperture
- data standardised in accordance with the standardisation rules (2004)

The approach was to initially develop IQI reference conditions for stable habitats exposed to low natural stress (fully marine, subtidal, fine sands/muds), with scope for adaptation once the classification tool had been tested within such environments.

IQI metric values were calculated for a range of data selected to include samples from sites subject to minimal degrees of anthropogenic pressure. The data were examined on a case-by-case basis using expert judgement to select metric values representative of reference conditions from the subset (Table 4.4). Metric values considered to be either anomalously high or erroneous values were removed from the process.

Table 4.4 $IQI_{v.IV}$ metric reference condition values (2006) established by UK and RoI competent authorities combining expert judgement and existing data

$IQI_{v.IV}$ metric	Reference condition value for fully marine subtidal fine sand/mud habitats (0.1 m^2, 1 mm sieve mesh)
Taxa number	68
1-(AMBI/7)	0.96
1-λ′	0.97

As such, a combination of existing data and expert judgement were adopted to establish reference conditions for the IQI for coastal water, fine depositional sediments.

The IQI standardisation rules (Prior et al. 2004) were revised in 2008 (see Chapter 2). To ensure the IQI assessments using the updated data treatment rules remained comparable with those made using the original data treatment rules, the metric reference condition values required updating (Table 4.5). The reference condition values were adjusted so that the average of the observed/reference condition values for each metric with the 2008 data treatment rules were consistent with those derived by applying the original 2004 data treatment rules and reference conditions.

Table 4.5 Revised IQI metric reference condition values for coastal sublittoral sand/mud habitats (0.1 m^2 grab samples, 1 mm sieve mesh) following adaptation of benthic data treatment rules (2008)

Metric	Reference condition value for fully marine subtidal fine sand/mud habitats (0.1 m^2, 1 mm sieve mesh): 2008 data treatment rules
Taxa number	78.6
1-(AMBI/7)	0.96
1-λ′	1.02 [1]

Note: [1] Reference values were adapted to ensure that the average values of the observed/reference constants were consistent for the data for the two different data treatment rules. The factor used to attain this consistency resulted in some revised reference conditions exceeding a value of one.

The revised reference conditions values, along with the relationship between the metrics and physicochemical conditions, sample collection and processing methodology, formed the basis for setting reference conditions for an expanded range of habitats (for example, low salinity, coarse/mixed substrata) and sampling methods (Figure 4.2).

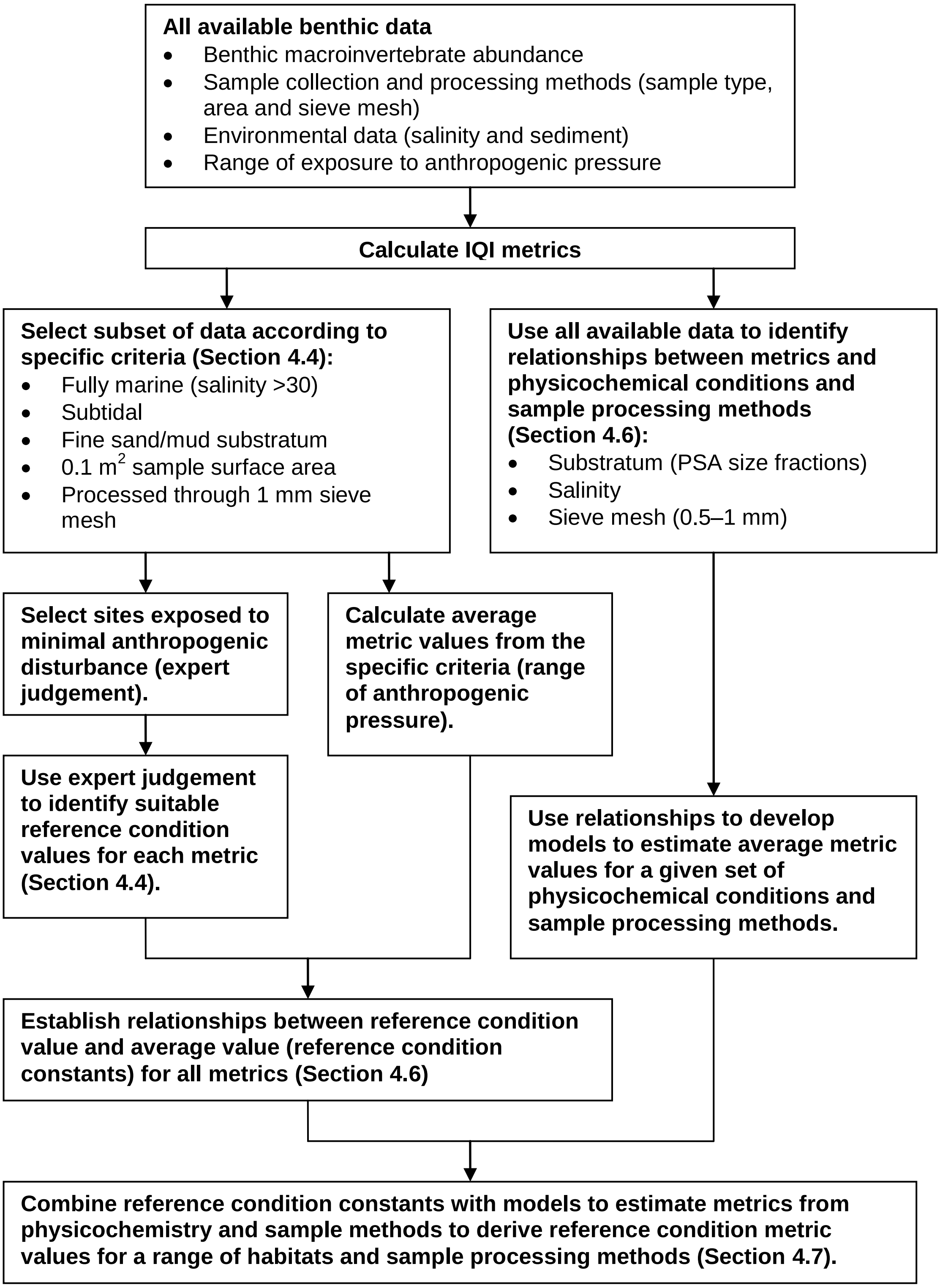


Figure 4.2 Overview of approach to setting IQI metric reference condition values

4.5 Defining habitats for reference conditions

Two approaches were considered in defining habitats for the purpose of setting reference conditions for the IQI metrics: ‘discrete’ and ‘continuous’ habitats. While it is acknowledged that the environment operates over continuous natural gradients, it is often divided arbitrarily according to physicochemical characteristics for classification and management purposes, for example, WFD typology (Figure 4.3).

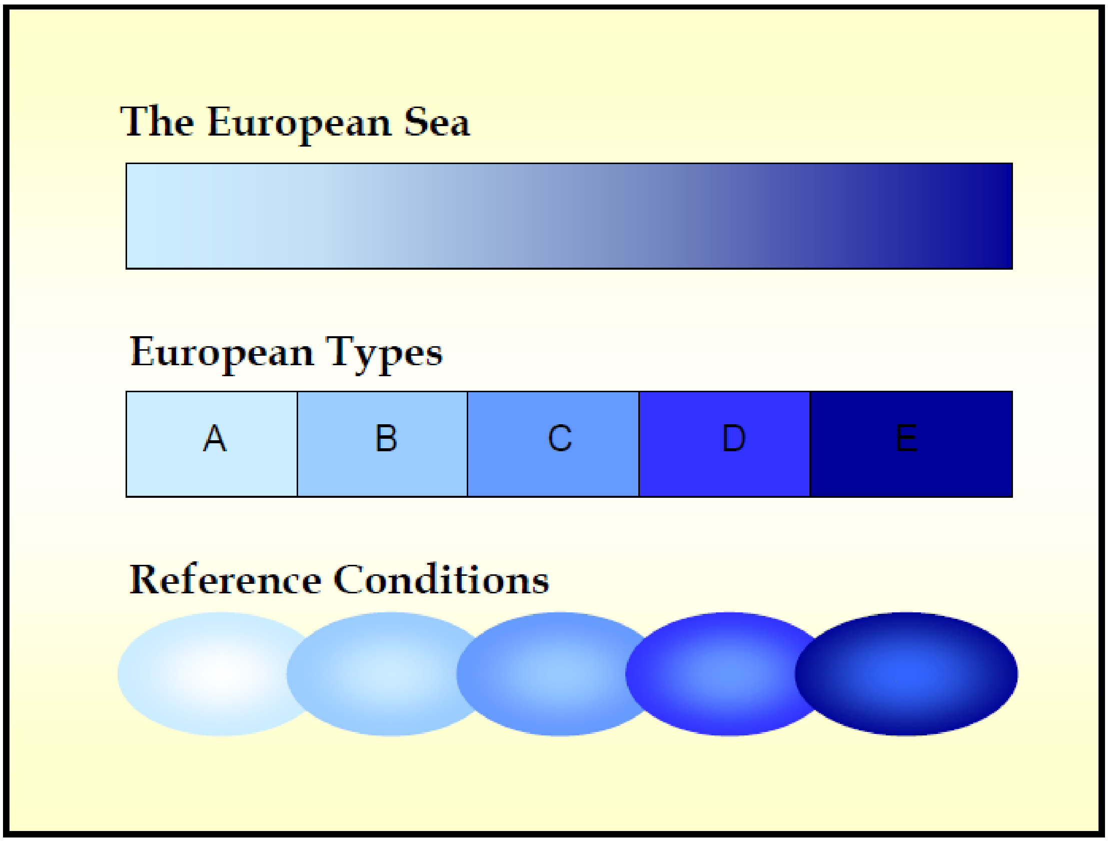


Figure 4.3 Relationship between all the seas in Europe (the European Sea), typology and type-specific reference conditions.

Notes: The European Sea is a continuum. Typology falsely compartmentalises this continuum into a number of physical types. The reference conditions for a specific water body type must then describe all possible natural variation within that type.
Modified from WFD CIS Guidance Document No. 5 (COAST 2003).

Water bodies are commonly further divided into separate habitats or zones such as the division of transitional water bodies into salinity zones according to the Venice system (Symposium on the Classification of Brackish Water 1959) (Figure 4.4).

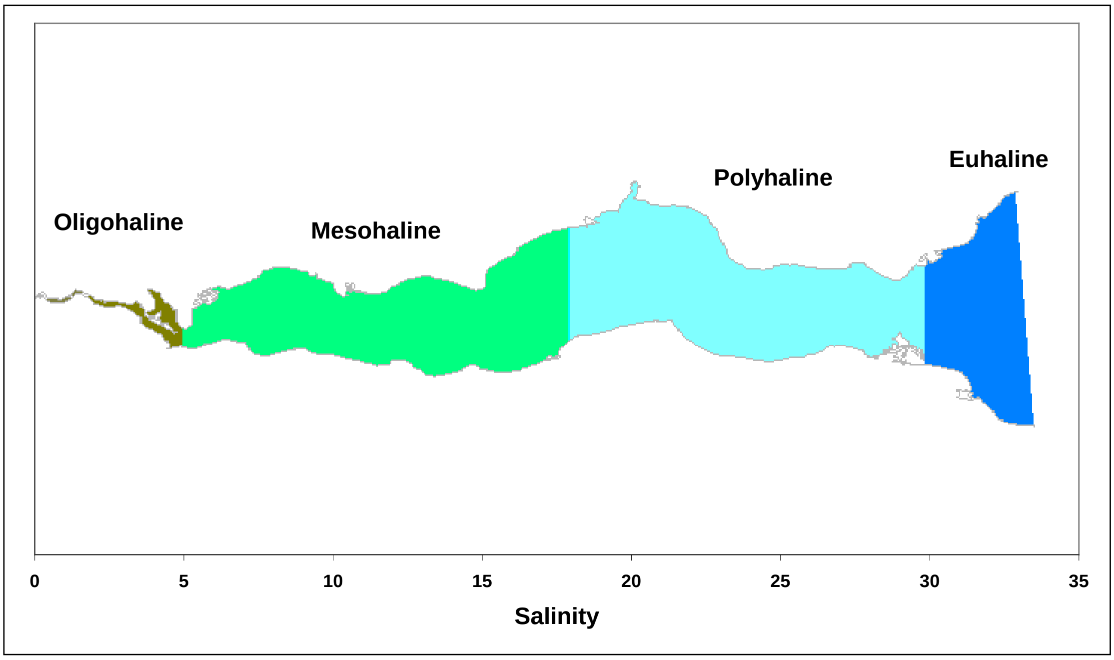


Figure 4.4 Conceptual diagram of the classification of salinity zones within a transitional water according to the Venice system (1959)

4.5.1 Discrete habitats

For classification systems that define habitats as discrete units, a site is categorised according to a combination of different environmental (for example, salinity, substratum, exposure, depth) and biological (in the case of biotope classification) parameters. Habitats divided and expressed as discrete units under such systems are referred to in this report as discrete habitats. Classification systems that are based on discrete habitats include:

- European Nature Information System (EUNIS) classification (Davies and Moss 1998, Davies et al. 2004)
- Marine Network Conservation Review (MNCR) Marine Habitat Classification (Connor et al. 1997a, 1997b, 2004)

Details of these two classification systems are given below. Approaches such as the EUNIS and MNCR systems provide a valuable means of summarising the characteristics of a site in an ecological or geomorphological context. Such systems for defining habitats have received widespread acceptance and reference in national and international environmental legislation.

EUNIS classification

The EUNIS habitat classification system is a pan-European system developed and managed by the European Topic Centre on Biological Diversity (http://eunis.eea.europa.eu). The system was developed to standardise habitat descriptions throughout Europe for habitat identification and includes those defined as natural, artificial, terrestrial, freshwater and marine. The system defines habitats as:

> ‘Plant and animal communities as the characterising elements of the biotic environment, together with abiotic factors operating together at a particular scale’.

EUNIS marine habitat types are classified at four hierarchical levels. Higher levels are divided according to physicochemical factors such as depth, exposure, association to seabed, hydrodynamic regime and substratum characteristics. Lower levels incorporate biotope descriptions according to certain individual taxa or taxonomic groups.

EUNIS is a widely adopted system used to assist habitat classification for purposes such as Natura 2000, which aims to establish an EU wide network of nature protection areas established under the Habitats Directive (92/43/EEC).

MNCR Marine Habitat Classification

The Marine Habitat Classification for Britain and Ireland (http://jncc.defra.gov.uk/marinehabitatclassification) was developed by the Joint Nature Conservation Council (JNCC) to define the marine habitats of the shores and seabed of Britain and Ireland. It adopts six levels of hierarchy, classified according to the following levels:

- environment
- broad habitats
- main habitats
- biotope complexes
- biotopes
- sub-biotopes

Since its original development, the MNCR classification system has adapted to align closely to the EUNIS system.

The MNCR classification system is used to define marine habitats for purposes such as establishing the conservation value of habitats for the identification of Marine Protected Areas (for example, Special Areas of Conservation and Special Protected Areas).

4.5.2 Continuous habitats

For the initial IQI development the EUNIS system provided an important method for identifying the habitat-specific data for the preliminary setting of the reference condition values (Prior et al. 2004). However, to improve the estimation of IQI reference condition values, an alternative means for describing marine habitats was proposed. Rather than define habitats qualitatively as discrete units, the approach attempted to express habitats across a continuous scale, according to measurable physicochemical parameters. These are referred to as continuous habitats.

4.5.3 Continuous versus discrete habitat classification

The use of continuous over discrete habitat classification has advantages and disadvantages as described below.

Advantages of using continuous habitats

Realistic representation of natural environmental gradients

In using discrete habitats the assumption is made that the similarity between benthic assemblages within a habitat is always greater than between different habitats. This may hold where habitats are differentiated by non-arbitrary features, such as the natural division between subtidal and intertidal, but does not necessarily hold where habitats are differentiated by arbitrary boundaries placed over continuous environmental gradients (for example, sediment particle size).

The advantage of using the continuous habitat approach in deriving reference conditions can be illustrated by observing the response of reference condition taxa number over a salinity gradient (Figure 4.5).

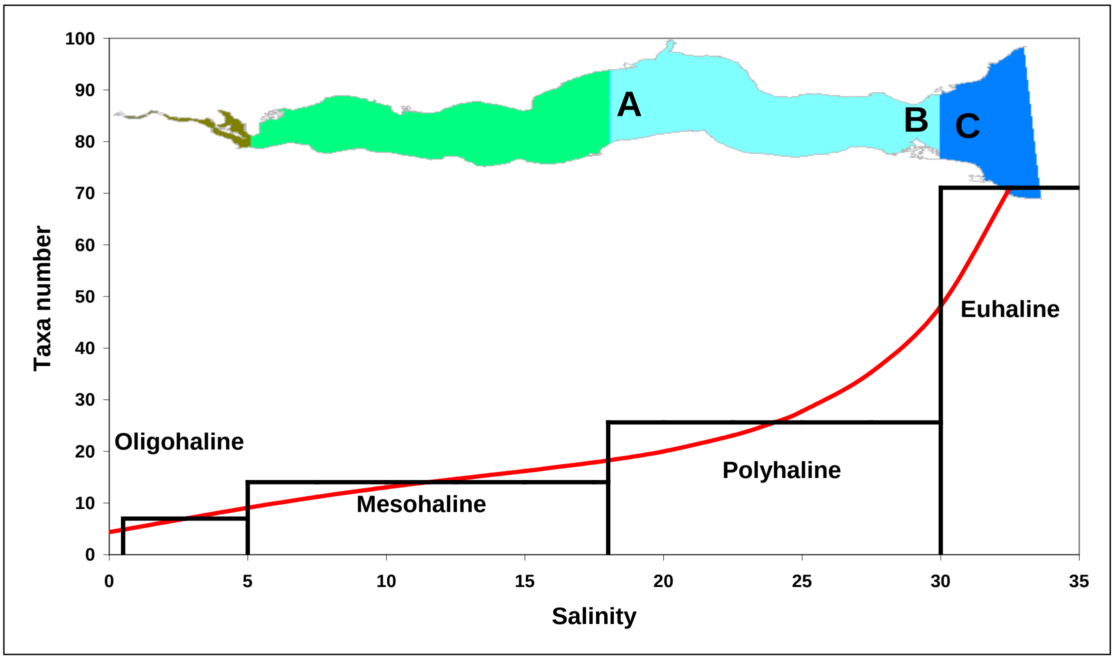


Figure 4.5 Reference condition taxa number values within a transitional water body for salinity zones classified according to the Venice system (discrete habitats) and operating over a continuum (continuous habitats)

Note: Taxa numbers based on draft reference conditions derived from Environment Agency 2007-2010 WFD surveillance monitoring data.

From the illustration shown in Figure 4.5, samples A and B are both classified as being from the same habitat (polyhaline) under the discrete habitat approach. As such, reference condition values at sites A and B would be identical despite an approximately two-fold increase in taxa number at site B within the habitat as indicated from the continuous plot of taxon number.

The consequence of assigning a single reference condition value to a habitat that experiences a high degree of within-habitat variation is that changes in metric values are likely to be interpreted as changing ecological condition, despite the changes being attributable to natural within-habitat bias. Alternatively, two samples experiencing similar physicochemical conditions but situated either side of a habitat boundary (samples B and C located within polyhaline and euhaline zones respectively) would be assigned markedly different reference condition values for taxa number (polyhaline = ~25, euhaline = ~70). This is despite the true difference being comparably minor as apparent from the continuous relationship between salinity and reference condition taxon numbers. Potentially, this could result in highly disparate EQR values between samples immediately either side of a habitat boundary, despite actual differences in anthropogenic disturbance being negligible.

Applying the continuous habitat approach aims to ensure that differences in reference condition values between two samples are proportionate to the differences in natural environmental conditions.

Ecological status boundaries (detailed in Chapter 5) are points on the EQR scale. As this scale is a ratio of the departure of observed values to the reference condition, deriving reference conditions according to a sliding scale means that the boundaries also operate over a sliding scale in terms of the absolute metric values (Figure 4.6).

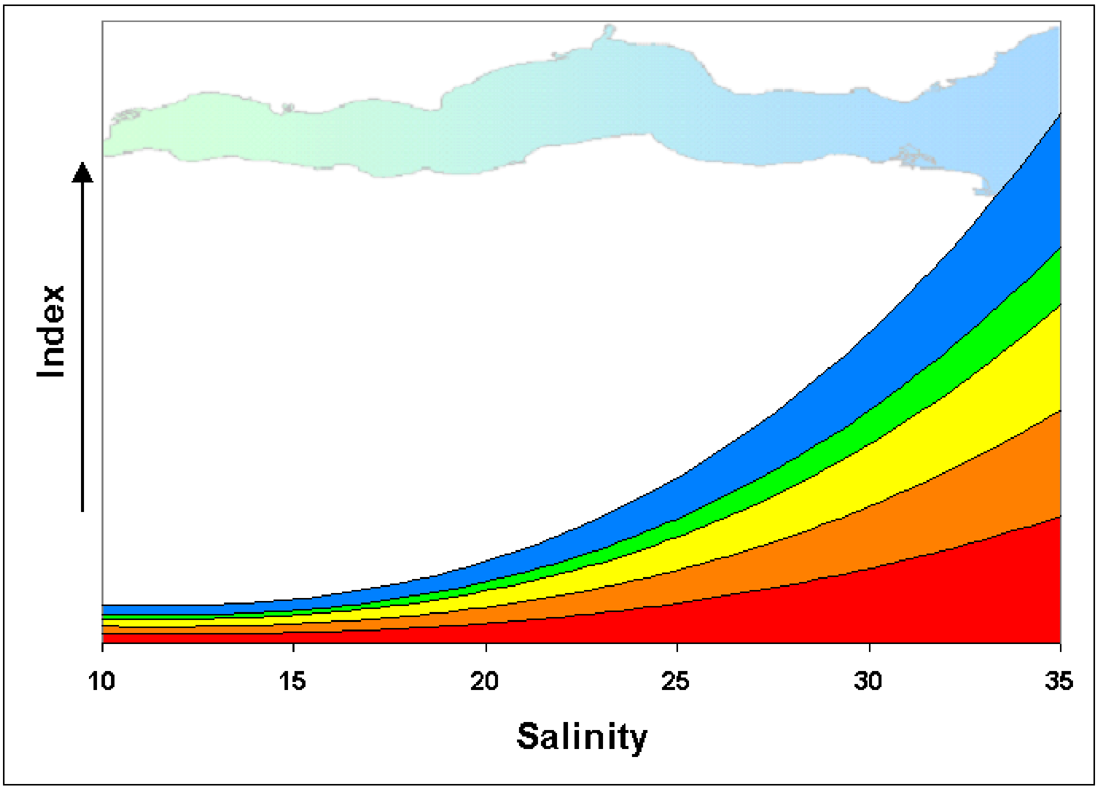


Figure 4.6 Changes in absolute values for a conceptual index with reference conditions operating over a continuum illustrating the how ecological status boundaries remain fixed relative to reference conditions

Increased objectivity in defining habitats

By using the continuous habitat approach, descriptions are based on quantitative physicochemical data that accompany the biological data, thus limiting the subjectivity introduced by using qualitative descriptions.

Discrete habitats may not be relevant to the IQI metrics

The rationale behind the categorisation of discrete habitats in existing classification systems may be of little relevance to the macrobenthic community as described under the WFD. For example, the division of habitats into EUNIS A5.2 or A5.3 depending on whether the silt/clay fraction is greater or less than 30% may not correspond to any meaningful change in taxa number, AMBI or 1-λ′ in the associated macrobenthic assemblage.

The continuous habitat approach is based on identified correlations between the physicochemical data (of those measured) and biological metrics, so reference condition values will only be affected by changes in physicochemical factors that correlate to changes in macrobenthic assemblages.

Disadvantages of using continuous habitats

Dependency on fully quantitative data

The continuous habitat approach requires the development of numerical relationships between physicochemical parameters and metric values. The effectiveness of this approach depends on the availability of standardised, quantitative environmental data. Where biological data are accompanied by environmental data, the data may be qualitative or, if quantitative, presented in a range of formats linked to the requirements of the study from which the data originate rather than to a standardised format.

A large amount of qualitative data and anecdotal evidence exists within the marine community. As this continuous habitat approach is reliant upon quantitative data, if there is no method of deriving quantitative values from qualitative data, this would make existing descriptive evidence unavailable for use in classification.

A potential solution to this disadvantage is described in Section 4.8.

Bias to reference condition values from physicochemical data collection methods

Physicochemical data (and therefore the values derived from them) may be biased as a result of the methods used in their collection. For example, spot salinity readings often taken alongside the benthic infauna are rarely sufficient in providing a complete record of the salinity regime (an important aspect when defining the macrobenthic community) at a given location. A habitat classification system based on qualitative data could allow for additional information (for example, anecdotal evidence) to be used to adapt reference conditions accordingly.

Difficulties in quantifying certain influential environmental characteristics

Additional habitat characteristics such as the dominance of maerl and the nature of shell material may be important factors influencing the IQI metrics. Such factors have not been quantifiable to date and therefore are not currently considered in estimating reference conditions. If such habitats are to be included for future WFD assessments, an objective approach to incorporating such factors needs to be established.

4.6 Approach to expanding IQI reference conditions

The initial reference condition values (Section 4.4) were set using data from a narrow range of discrete habitats (EUNIS type A5.2 and A5.3 habitats: fully marine subtidal sand and mud respectively) and a single sampling collection and processing method (0.1 m^2 grab, 1 mm sieve mesh). To maximise the applicability of the IQI in classifying transitional and coastal waters, these reference condition values had to be adapted to consider a wider range of habitats and sampling methodologies. This was done by using the available physicochemical data to model reference condition values across salinity and sediment continua.

The first step in the process involved identifying the numerical relationship between each metric and the measured environmental (physicochemical) parameters. Once these relationships were identified, an expected metric value could be estimated for any given combination of physicochemical conditions. This estimated metric value is referred to in this report as E(X).

The second step in the process was to identify the relationship between the average metric values estimated from the physicochemistry and the expected reference condition values. The approach is based on the philosophy that the relationship between E(X) and the reference condition value should be constant throughout all habitats and sampling methods.

Details of the process are described as follows.

4.6.1 Step 1: Relating metrics to physicochemical parameters

The approach to identifying the relationship between the IQI metrics and the physicochemical parameters adopts the principle that the total variability observed in each metric throughout the available data is a function of:

- natural environmental conditions
- anthropogenic pressures
- random (sampling/measurement and ecological interaction) error

This approach can be expressed by Classical test theory (further described in Chapter 6):

$$X = T + e_s + e_r \qquad \text{Equation 4.2}$$

where:

X = observed value (metric)

T = true value (response to anthropogenic pressure)

e_s = systematic error (bias from natural pressures)

e_r = random error (sampling/measurement and ecological interaction)

In the current context, the principle assumes that a true response to bias from anthropogenic pressure (T) exists which would be evident from the metrics if no error existed.

Benthic communities from naturally stressed areas share many characteristics of areas suffering from anthropogenic stress (Elliott and Quintino 2007). On the basis that the metrics are therefore likely to respond to natural pressures in a comparable way to anthropogenic pressure, part of the total variability in the observed metrics should correspond to the measured supporting physicochemical parameters (the systematic error (natural bias) or e_s in Equation 4.2). It is therefore possible to use the relationships between the metrics and the supporting physicochemical data to derive models to estimate metric values that occur as a result of the natural conditions alone (that is,, the measured systematic error, or e_s). However, the data used in developing the models are subject to anthropogenic disturbance and as such the values estimated from the models do not directly equate to systematic error (e_s), but instead represent systematic error with a function of anthropogenic pressure (T) and random sampling/measurement and ecological interaction error (e_r). The data used to develop the models contain biological and environmental data exposed to varying degrees of anthropogenic pressure and have not been selected purely as data representative of reference conditions.

For the approach to modelling metric values from the data, the following assumptions were applied.

- The average extent of bias resulting from anthropogenic pressure (T) is constant throughout the data.
- All natural drivers influential to the variability of the metrics have been correctly captured within the models.
- Random sampling/measurement and ecological interaction error (e_r) is assumed to have a mean of zero (exogeneity).

Based on these assumptions, the metric values estimated from the available physicochemical data using the models (termed E(X)) are assumed to correlate to the systematic error (e_s). It is acknowledged that, in reality, these assumptions are unlikely to hold (see discussion).

In a hypothetical scenario where without any anthropogenic disturbance and random error, if a metric were to correspond exactly to a measurable natural parameter (that is, the systematic error/natural bias (e_s) was fully understood), measurements of the parameter could be used to precisely estimate the metric E(X) (Figure 4.7a). However, estimates of E(X) from the physicochemical data are prone to the effects of random error in terms of their departure from values expected from natural systematic bias alone and anthropogenic pressure (illustrated in Figure 4.7b and Figure 4.7c respectively). The effects of anthropogenic pressure and random error generate what is expected in terms of the variability within the real data (Figure 4.7d).

If the extent of systematic error was known and random error was zero, the true effect of anthropogenic pressure could be accurately calculated from an observed value of the metric. This is not possible in reality. However, by using the available physicochemical data to estimate a degree of systematic bias for each metric, the approach enables a closer estimation of the extent of anthropogenic pressure and random error. In theory, if all systematic bias attributable to natural pressure could be predicted and their effect on the observed metric removed, the remaining variability would correspond directly to anthropogenic pressure and random error, that is, $X - e_s = T + e_r$ (Figure 4.8).

To further attain a closer estimation of the effect of anthropogenic pressure, the effect of random error should also be reduced (it is assumed in the process to have a mean of zero). Again, in theory, reducing random error to zero would enable an exact calculation of the effect of anthropogenic pressure from the observed metric value if systematic bias was also known: $X - e_s = T + 0$, and therefore $X - e_s = T$.

While it is not possible to eliminate random error (e_r), it can be reduced by adopting:

- appropriate sampling design;
- sample collection, processing and analysis controls (for example, operational instructions, best practice guidance, analytical quality control procedures)
- data treatment protocols

In addition, variability within e_r also results from ecological processes such as predator/prey interactions, competition and alteration by natural disturbance events (for example, storms, unusual levels of freshwater flow). Such aspects are an accepted part of ecological sampling and contribute to the overall error that must be addressed in deriving an appropriate sampling programme.

Where a metric displays a monotonic relationship with pressure, the departure of the observed values from those estimated from a selection of natural physicochemical conditions are likely to correlate with pressures not accommodated within the predictive models, as indicated by the direction of increasing pressure in Figure 4.8. By using systematic bias that correlates to natural pressures for the metric estimations, anthropogenic pressures are expected to have an increased influence on the departure of observed from estimated metrics (that is, the direction of increasing pressure should more closely correspond to anthropogenic pressure). This can be translated into the classical test theory model (Equation 4.2) whereby observed values greater than estimated values ($X > E(X)$) are expected to correspond to lower degrees of pressure (T) than when observed values are less than estimated values ($X < E(X)$). Likewise, observed values lower than estimated are expected to correspond to higher degrees of pressure.

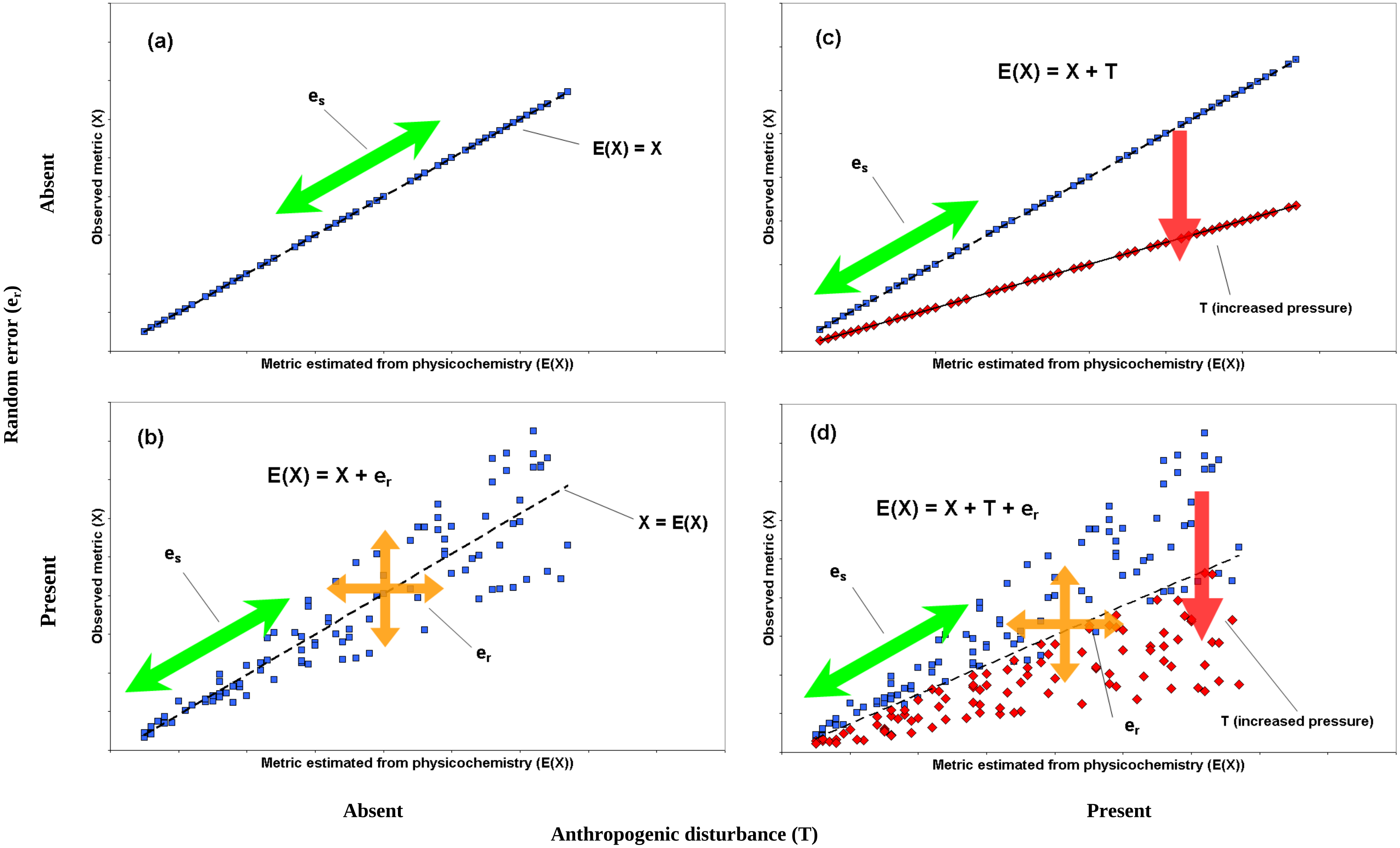


Figure 4.7 The principles of using physicochemical data to estimate the IQI metrics (E(X)) illustrating where (a) observed values are driven by environmental conditions without error, (b) observed values are driven by environmental conditions and subject to random, sampling and measurement error, (c) observed values are driven by environmental conditions without error and the effect of anthropogenic disturbance, and (d) observed values are driven by environmental conditions with error and the effect of anthropogenic disturbance

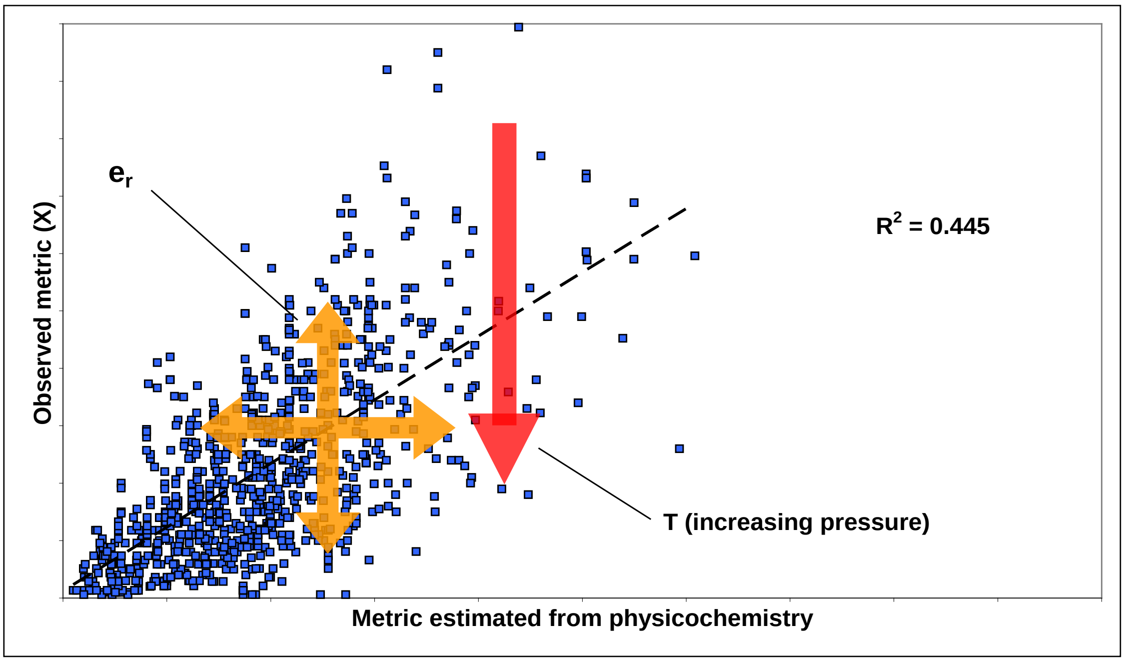


Figure 4.8 Example of observed metric values (X) versus values expected as a result of systematic bias from natural pressures (e_s) indicating how departure from X = E(X) are expected to be influenced by anthropogenic pressure (T) and random sample and measurement error (e_r)

Note: In general, decreases in observed metric values are expected to correspond to increasing anthropogenic pressure.

The Pearson correlation coefficient (R^2) value reflects the extent to which the variability in the observed metrics can be explained by the physicochemical data (the estimated metrics). For illustration purposes, it is assumed that all systematic bias from natural pressure is incorporated within the estimation of e_S. However, this is not the case as the remaining variability within T + e_r (that is, the departure of the data from X = E(X)) will not only result from anthropogenic influences and random error, but also include variability attributable to unrecorded natural environmental factors and insufficiencies in how those factors recorded are included in the model (the relationship between parameter and metric are not accurately represented in the model).

The metrics were related to the physicochemical conditions using multiple regression. Physicochemical data were used as the predictor variables in estimating the metric response variable (E(X)):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p \qquad \text{Equation 4.3}$$

where:

Y = dependent/response variable (metric E(X))

X = independent/predictor variable (physicochemical parameter)

β_0 = constant coefficient

β = predictor variable coefficient

By incorporating second-order (squared) transformations of the physicochemical parameters, the models become quadratic regressions acknowledging that the relationships between the dependent variable (metric (E(X)) and the independent variables (physicochemical parameters) may be curvilinear.

Data used to develop the models

The models used to estimate metric E(X) values were developed using all available data from across as broad a range of physicochemical conditions as possible, with exposure to varying degrees of anthropogenic pressures. For the purpose of the models, it was assumed that the degree of anthropogenic pressure is constant across the full range of physicochemical conditions in the data, that is, there is no correlation between anthropogenic pressure and any of the physicochemical conditions (this is not expected to be the true situation – see discussion in Section 4.9). Modelling the reference condition values on data isolated from pre-determined ‘high status’ reference condition sites would have the adverse effect of reducing the objectivity of the models (as reference condition sites would need identifying using expert judgement).

The physicochemical variables used for the metric E(X) model development were:

- PSA % <63μm
- PSA % 63 ≤125μm
- PSA % 125 ≤250μm
- PSA % 250 ≤500μm
- PSA % 500 ≤1,000μm
- PSA % 1,000 ≤2,000μm
- PSA % 2,000 ≤4,000μm
- PSA % 4,000 ≤8,000μm
- PSA % ≥8,000μm
- mean salinity*
- salinity standard deviation*

*To ensure that the salinity values assigned to each benthic station were closely representative of the salinity at the site, data from <1,000m of each benthic station were used from a range of datasets.

Note: The physicochemical variables listed above are those used in the approach to deriving the initial reference condition models. The parameters included are to be revised as additional data becomes available. It is recommended that the continued development of reference conditions explores the use of variables not listed above that may be potentially important factors in estimating metric values from environmental data (for example, upper/lower salinity percentiles, sediment statistics).

The IQI metric E(X) models were developed using data from the CSEMP and WFD surveillance (see Section 6.7) programmes. These biological datasets have the following advantages.

- They contain matched quantitative physicochemical supporting data.
- They include data from a broad spatial range (national level).

- They include data from varying degrees of anthropogenic pressure
- They follow standardised sampling methods.
- They are subjected to standardised AQC procedures through the NMBAQC scheme.

The majority of available supporting salinity data were based on spot sample sampling. Due to the variation in salinity at a single location resulting from changes over the tidal cycle, seasonal variation due to freshwater input and so on, spot sample salinity data provide a highly limited description of the characteristics of a site.

Salinity values used within the initial regression model were therefore based on the average and standard deviation of multiple salinity spot samples within a given proximity of the biology sites. Samples were excluded from the analysis if they did not meet the standardised sampling criteria. Based on the dominant available data (paired biology and physicochemistry), reference condition models were established for 0.1 m^2 subtidal grab and 3×0.01 m^2 intertidal core (pooled samples) sample methods.

Before modelling the metrics, it was necessary to transform the metrics and the physicochemical data. For the purpose of confidence interval and hypothesis testing, regression analysis requires assumptions of normality and homoscedasticity (homogeneity of variance) for the dependent variables (metrics). While assumptions of normality and homoscedastity are not crucial for estimating the regression parameters, the metrics were transformed to achieve closer approximations to these assumptions (Table 4.6).

Table 4.6 Distribution of metric data with associated transformation

Metric (dependent variable)	Data distribution	Appropriate transformation
Taxa number	Left-skewed	Log(x+1)
1-(AMBI/7)	Binomial	Arcsine
1-λ′	Binomial	Arcsine

Accommodating sieve size fractions

In order to utilise data from different sieve mesh sizes (to maximise the amount of data for each regression model), the 0.5 and 1 mm sieve mesh data needed to be incorporated into a single dataset. The metrics from the 1 mm fraction were utilised directly, but the metrics from the 0.5 mm fraction were adjusted to estimate the metrics expected at the 1 mm equivalent. The relationships between the transformed metrics from the 0.5 and 1 mm fractions can be expressed as linear regressions as follows:

$$Y = \alpha + \beta X \qquad \text{Equation 4.4}$$

where:

Y = transformed metric for 1 mm fraction

α = intercept

β = constant

X = transformed metric for 0.5 mm fraction

The regression coefficients used to adjust the 0.5 mm metrics to approximate 1 mm equivalents for grab and core sample methods are presented in Table 4.7. The relationships are illustrated in Figures 4.9 to 4.14.

Table 4.7 Regression coefficients for the conversion of transformed 0.5 mm metric values to 1 mm equivalent

Method	Metric	Transformation	Intercept (α)	Constant (β)
0.1 m^2 grab	Taxa number	Log(x+1)	-0.303	1.002
	1-(AMBI/7)	Arcsine	0.044	0.996
	1-λ′	Arcsine	0.115	0.811
3 × 0.01 m^2 core	Taxa number	Log(x+1)	-0.154	0.908
	1-(AMBI/7)	Arcsine	0.024	1.141
	1-λ′	Arcsine	0.311	0.567

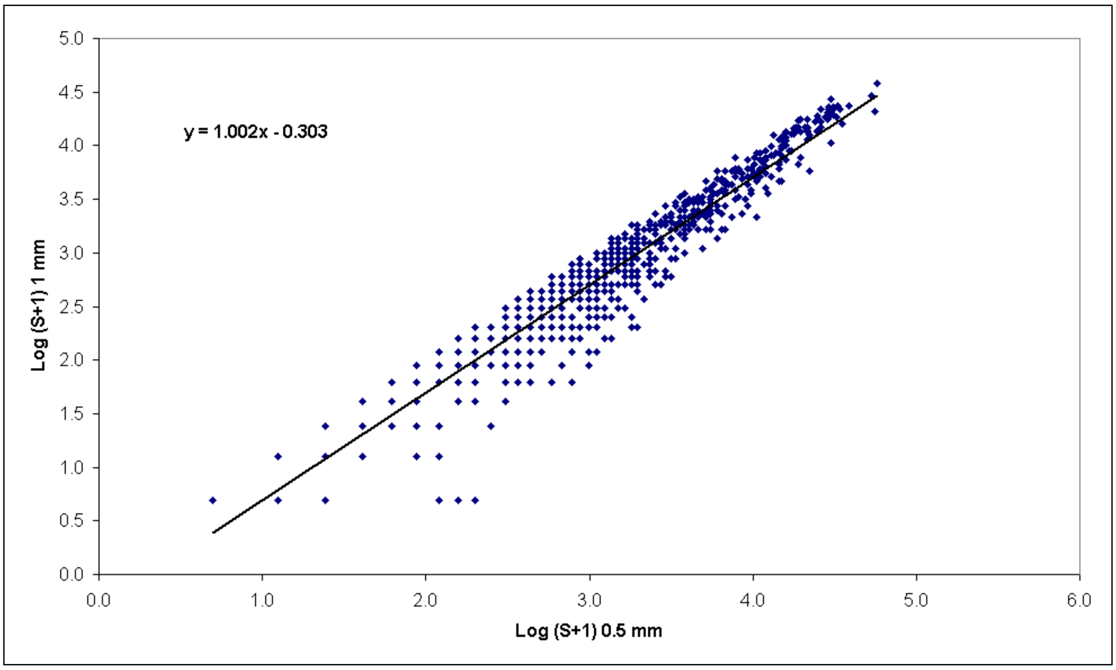


Figure 4.9 Transformed taxa number (log(x+1)) for 0.5 mm sieve mesh versus corresponding transformed taxa number for 1 mm sieve mesh for 0.1 m^2 subtidal grab data (n = 866, source: CSEMP)

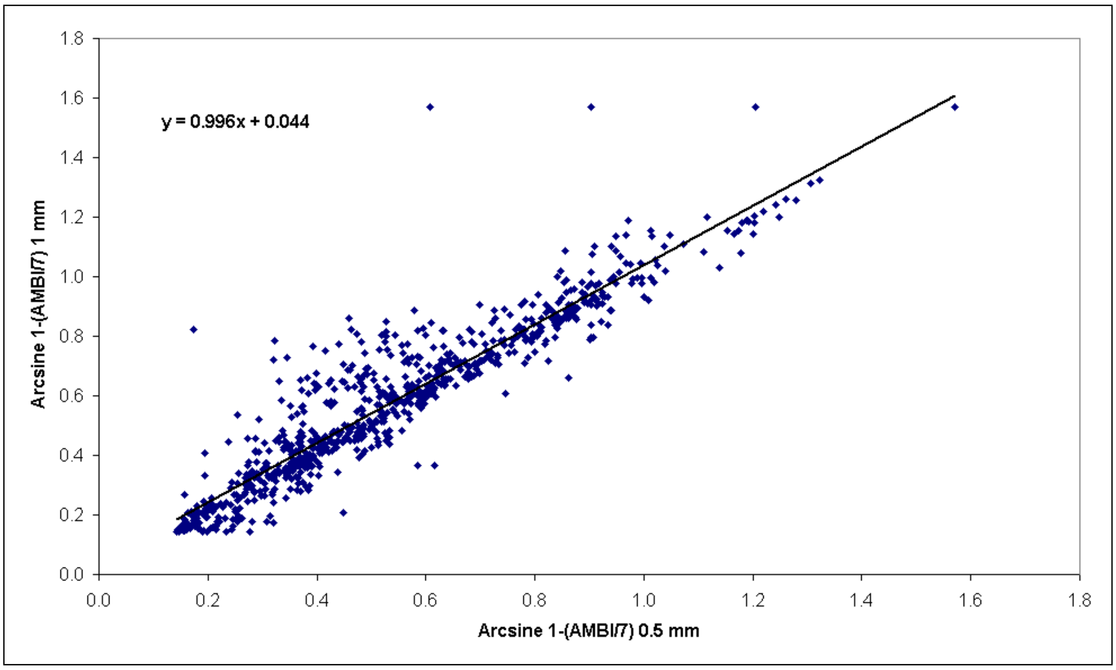


Figure 4.10 Transformed 1-(AMBI/7) (arcsine) for 0.5 mm sieve mesh versus corresponding transformed 1-(AMBI/7) for 1 mm sieve mesh for 0.1 m^2 subtidal grab data (n = 866, source: CSEMP)

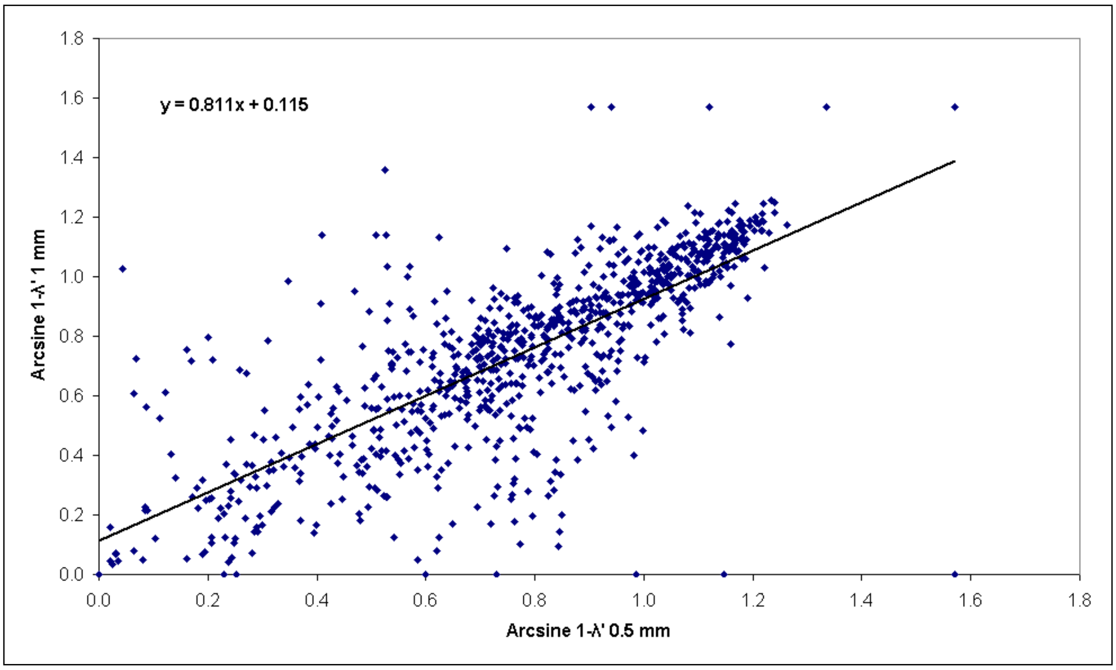


Figure 4.11 Transformed 1-λ′ (arcsine) for 0.5 mm sieve mesh versus corresponding transformed 1-λ′ for 1 mm sieve mesh for 0.1 m^2 subtidal grab data (n = 866, source: CSEMP)

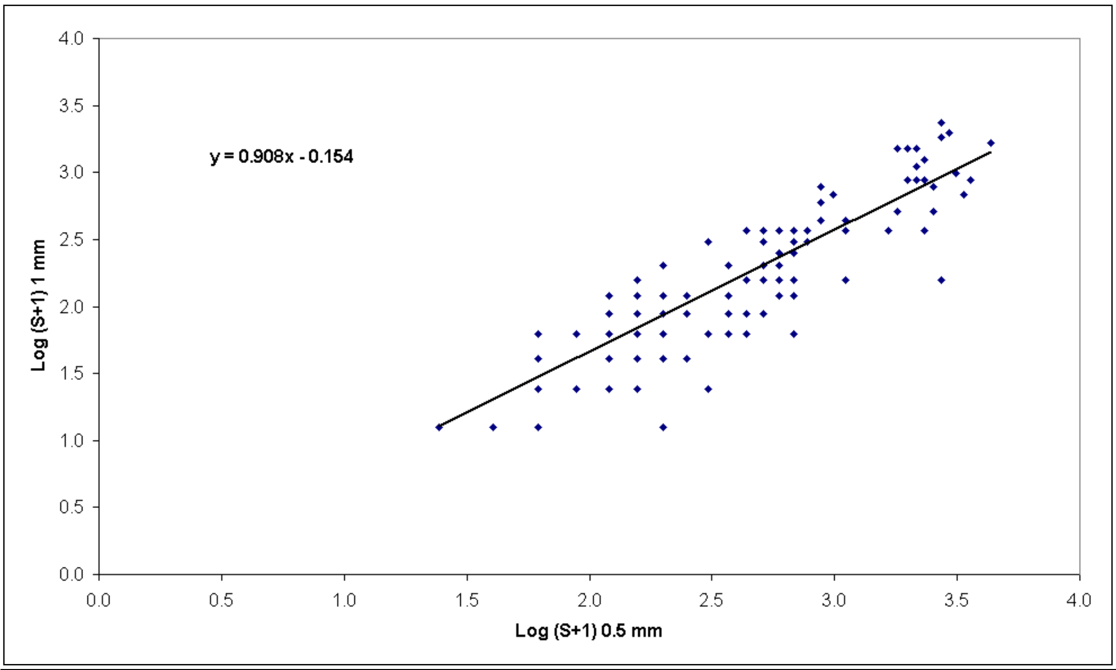


Figure 4.12 Transformed taxa number (log(x+1)) for 0.5 mm sieve mesh versus corresponding transformed taxa number for 1 mm sieve mesh for 0.01 m^2 intertidal core data (n = 135, source: CSEMP)

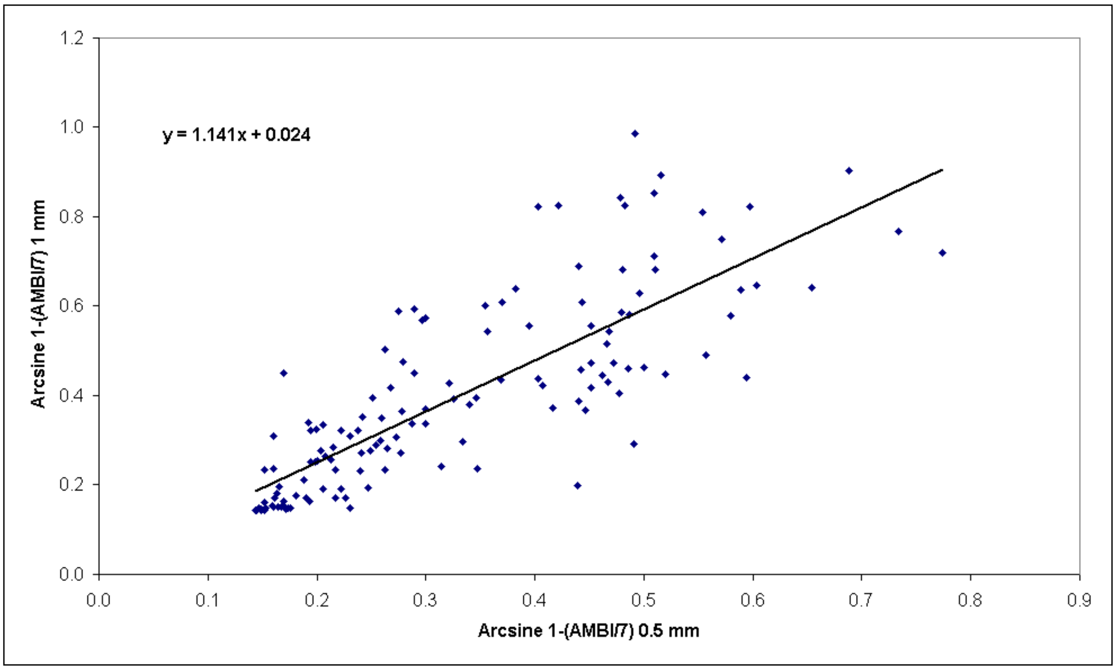


Figure 4.13 Transformed 1-(AMBI/7) (arcsine) for 0.5 mm sieve mesh versus corresponding transformed 1-(AMBI/7) for 1 mm sieve mesh for 0.01 m^2 intertidal core data (n = 135, source: CSEMP)

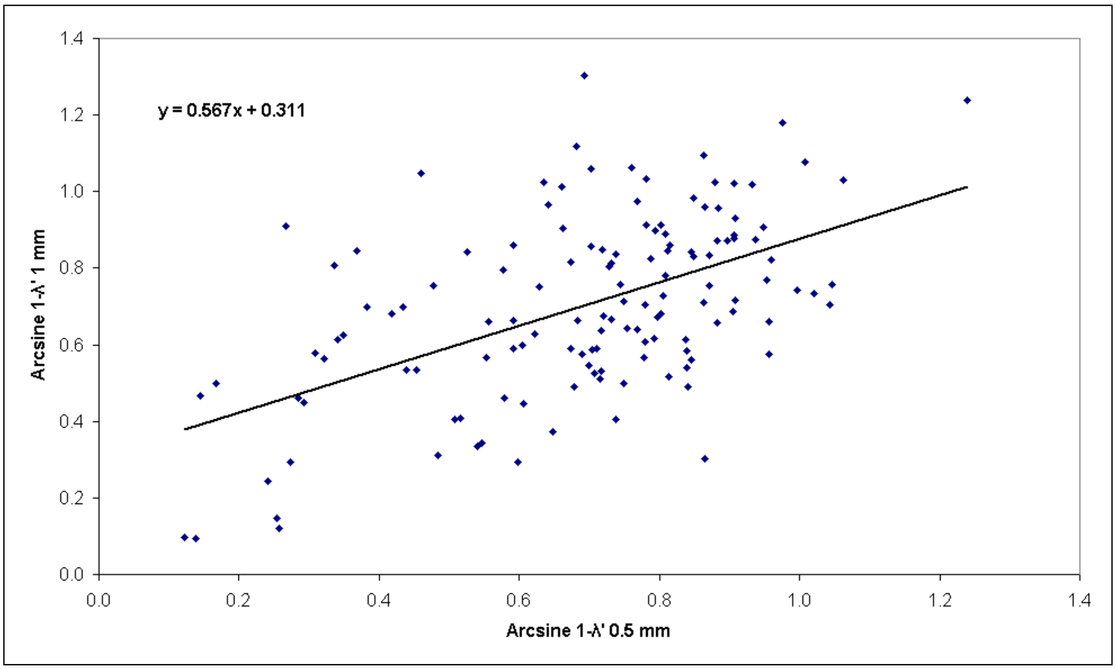


Figure 4.14 Transformed 1-λ′ (arcsine) for 0.5 mm sieve mesh versus corresponding transformed 1-λ′ for 1 mm sieve mesh for 0.01 m^2 intertidal core data (n = 135, source: CSEMP)

An alternative approach to enable pooling of 0.5 and 1 mm data would be to include mesh size as an additional predictor variable within the regression model. This approach was rejected on the basis the sieve mesh used generally corresponded to monitoring of transitional (0.5 mm) and coastal (1 mm) sites. Elevated levels of salinity variability experienced within transitional waters (and therefore the majority of 0.5 mm data in the analysed dataset) were expected to correspond to reducing metrics values. If data from both size fractions were incorporated into the regression formula unaltered, it would not be possible to differentiate the bias caused to the index values resulting from changing salinity from the bias resulting from a change in sieve size.

Physicochemical data transformation

The physicochemical data were transformed to potentially further reduce the curvilinearity of the relationships between predictor and response variables. Salinity data (average and standard deviation) were log(x+1) transformed and particle size fractions arcsine transformed (percentages divided by 100 to be expressed as proportions for arcsine transformation).

Regression analysis (physicochemistry versus metrics)

Regression analysis was performed between each transformed metric (response variable) and the range of transformed physicochemical parameters (predictor variables) using Minitab© statistical software. As quadratic regression was used, the square of each transformed physicochemical variable was included in the analysis.

Subset analysis using Mallows C_p was initially undertaken on the full set of physicochemical predictor variables. This was to reduce the effect of overfitting, whereby variables that are likely to correspond only to random error in the model data are excluded, thus improving the ability of the models for predicting reference condition values in additional datasets. The analysis was repeated for each metric for subtidal 0.1 m^2 grab and intertidal 3 $\times$ 0.01 m^2 core (pooled) data, both processed using a 1 mm sieve mesh.

The extent to which the variability within the physicochemical data can be used to explain variability in the (observed) metric data is expressed by the regression analysis Pearson R^2 values (Table 4.8).

Table 4.8 Pearson correlation coefficients (R^2) between observed metric values and expected values (E(X)) estimated from physicochemical data using the regression models

Sample method	Metric	Pearson correlation (R^2)
0.1 m^2 grab	Log(taxa number+1)	44.4%
	Arcsine(1-(AMBI/7))	44.6%
	Arcsine(1- λ′)	28.9%
3 $\times$ 0.01 m^2 core	Log(taxa number+1)	52.7%
	Arcsine(1-(AMBI/7))	45.2%
	Arcsine(1-λ′)	10.9%

The regression formula describing the relationship between the observed metrics and physicochemical data was then used to estimate metric E(X) values at the sample level based on the sample specific physicochemical conditions. The extent to which the natural physicochemistry (expressed by the E(X) values) explains the variability of the observed metric values can be observed for all metrics in Figures 4.15 to 4.20.

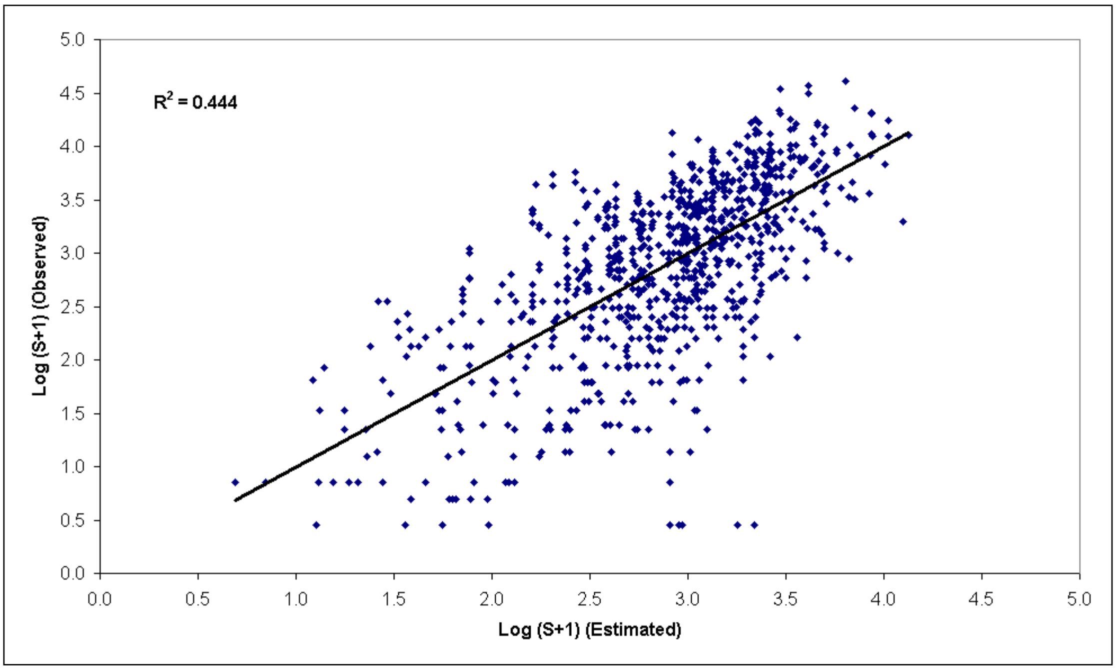


Figure 4.15 Observed transformed taxa number versus estimated transformed taxa number (0.1 m^2 grab samples)

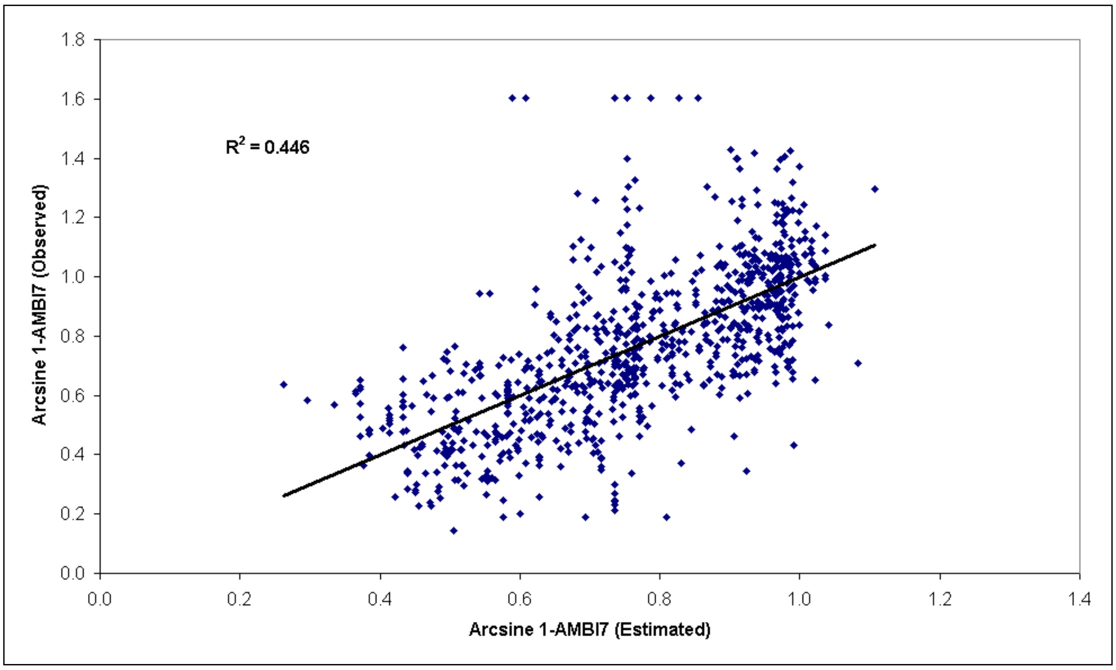


Figure 4.16 Observed transformed 1-(AMBI/7) versus estimated transformed 1-(AMBI/7) (0.1 m^2 grab samples)

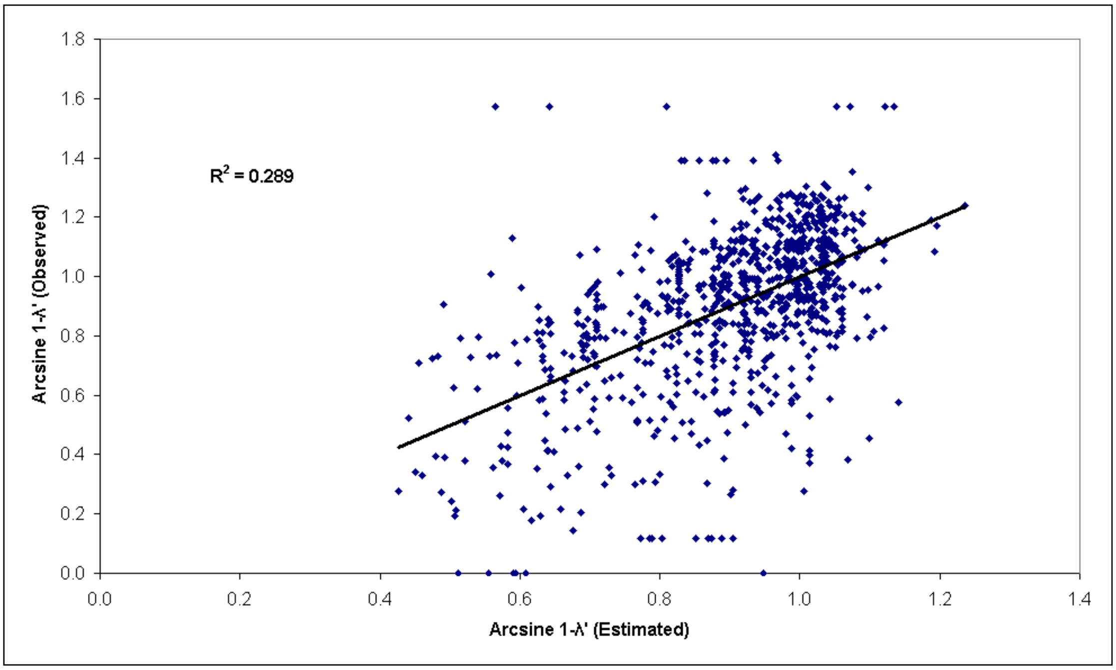


Figure 4.17 Observed transformed 1-λ′ versus estimated transformed 1-λ′ (0.1 m^2 grab samples)

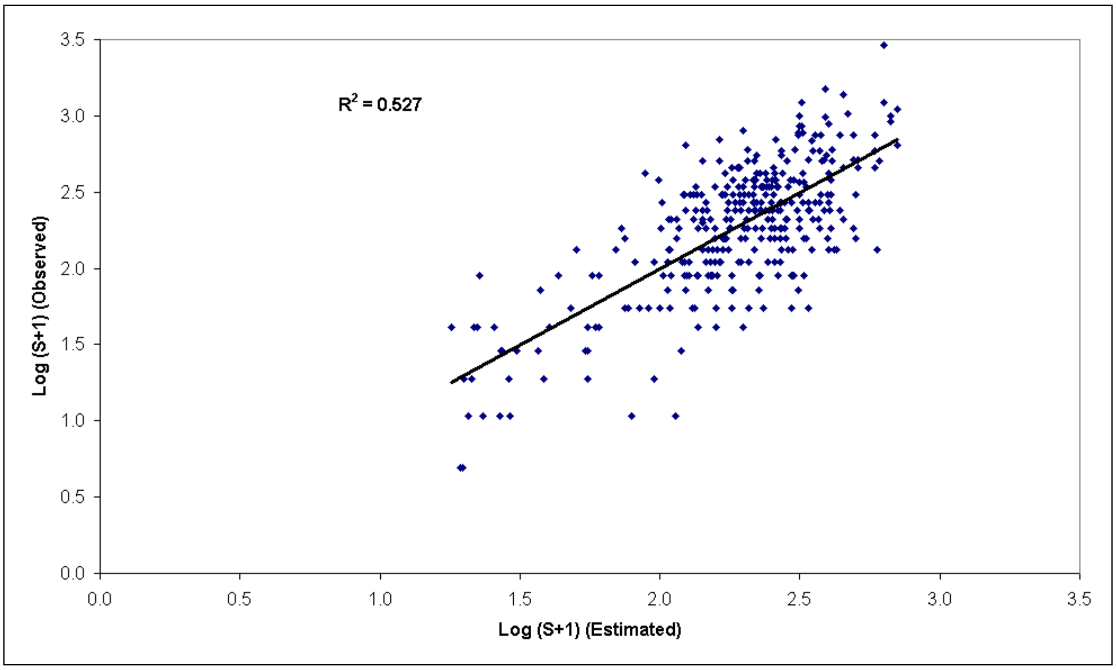


Figure 4.18 Observed transformed taxa number versus estimated transformed taxa number (3 × 0.01 m² core samples)

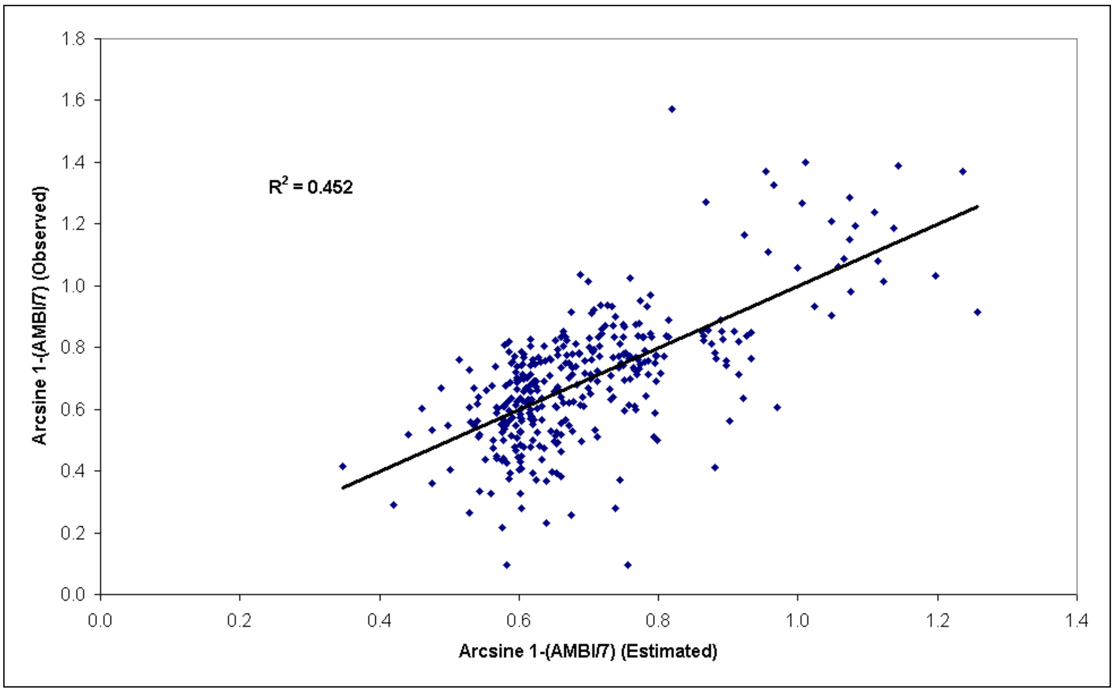


Figure 4.19 Observed transformed 1-(AMBI/7) versus estimated transformed 1-(AMBI/7) (3 × 0.01 m² core samples)

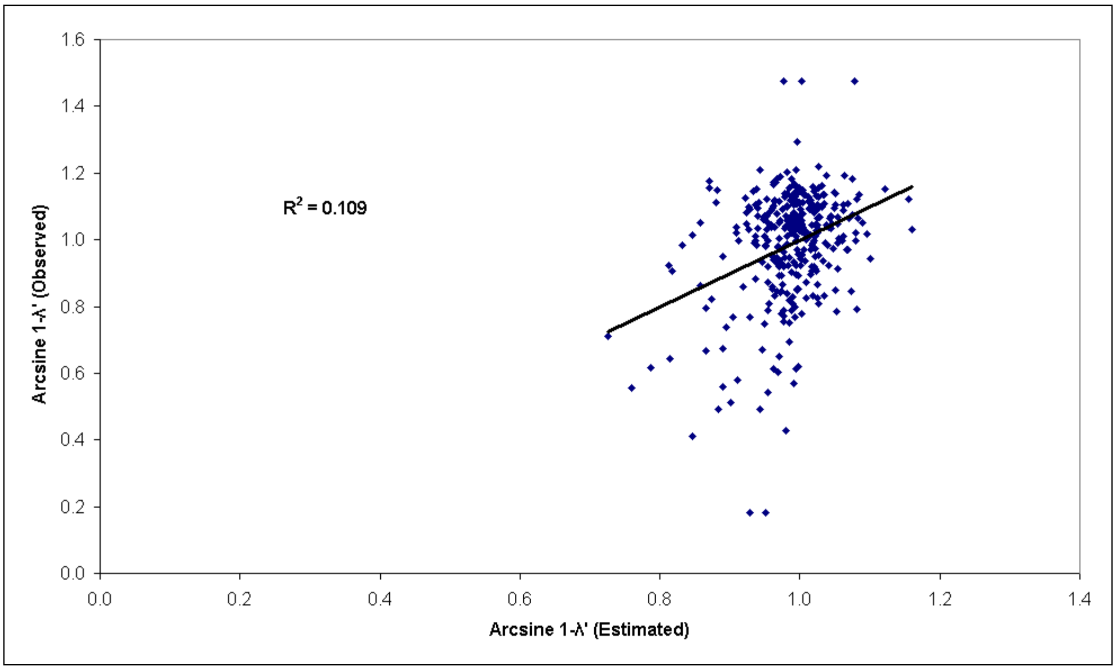


Figure 4.20 Observed transformed 1-λ′ versus estimated transformed 1-λ′ (3 × 0.01 m^2 core samples)

The extent to which the variability of the transformed observed metrics can be ‘explained’ by the physicochemical data ranges from a Pearson R^2 of 10.9% for 3 × 0.01 m^2 core arcsine(1-λ′) data to an R^2 of 52.7% for 3 × 0.01 m^2 core log(taxon number+1) data. Where the unexplained variability is high (low correlation between observed and estimated metric values), this indicates that the environmental data used to estimate the metrics accounts for low degrees of the overall variability of the metric. The remaining variability not explained by the measured physicochemistry potentially results from several factors such as additional non-recorded physicochemical factors (for example, temperature, depth, hydrodynamic aspects) and the response to anthropogenic influences.

The total variability observed in the data is the sum of the signal (anthropogenic disturbance) and the natural variability or ‘noise’ (systematic bias and sampling and measurement error). To increase the ability of the IQI to detect an anthropogenic disturbance signal it is important that the systematic bias is understood and that the sampling and measurement error is minimised. By accounting for as much systematic bias or natural variability as possible in the reference condition models, the proportion of the remaining variability explainable by anthropogenic disturbance is increased, therefore increasing the effectiveness of the IQI in detecting such anthropogenic disturbance. Variability in the IQI is the focus of Chapter 6.

4.6.2 Step 2: Relating estimated metrics to reference conditions

The second step in the process was to identify the relationship between the metric E(X) values estimated from the physicochemistry and those exposed to minimal degrees of anthropogenic pressure (reference conditions or X_{Ref}), that is, the departure of X_{Ref} from E(X). Any given EQR should represent a consistent departure of observed metric values from those at reference condition, independent of the physicochemical conditions and the sample collection and processing methodology.

To ensure this principle is applied, the extent to which reference condition metric values exceed the estimated metric values (that is, departure of X_{Ref} from E(X)) should be proportional across the range of physicochemical conditions and sample methodologies (Figure 4.21). On this basis, reference condition values were developed to consistently follow the relationship between metrics and physicochemical conditions, being a fixed proportion increase over the value expected according to the natural conditions alone. Under this approach, the status boundaries therefore also consistently follow the relationship between the metrics and physicochemical conditions. This constant value to be applied to calculate reference conditions from metric E(X) values is being termed the 'reference condition constant':

$$\frac{X_{\text{Ref}}}{E(X)} = \text{Reference condition constant}$$

Equation 4.5

By multiplying the estimated metric E(X) values from across the range of physicochemical parameters by the reference condition constant, reference condition values for an expanded range of habitats were established.

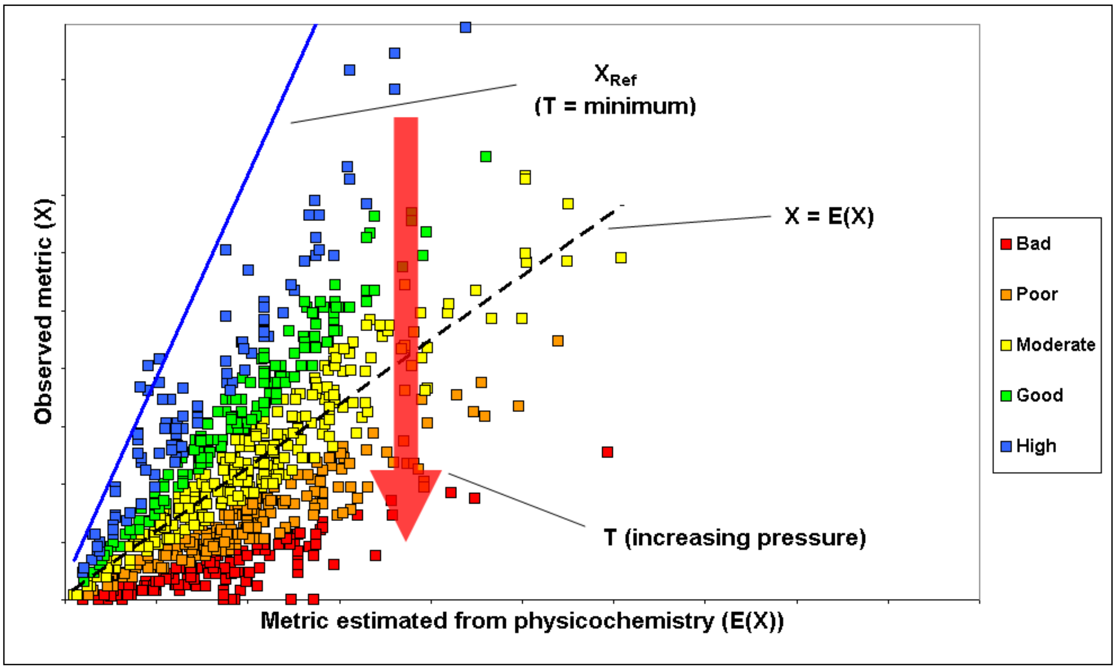


Figure 4.21 Hypothetical example of observed metric values (X) versus metric values estimated from physicochemistry (E(X)) indicating relationship between reference conditions (X_{Ref}) and metric E(X), expected influence of anthropogenic pressure (T) and indicative status classes in relation to X and metric E(X) values

Once the metric E(X) values had been calculated from the available physicochemical data, the relationship between E(X) and the reference conditions was established. Average physicochemical parameters[9] for EUNIS A5.2/A5.3 were calculated from the WFD and CSEMP data (Table 4.9) and applied to the regression formula to estimate metric E(X) values for the conditions under which the original reference conditions were derived.

[9] Weighted average value for each sediment size fraction and average salinity (coastal water body data)

Table 4.9 Weighted average PSA size fractions and salinity values for EUNIS A5.2/A5.3 habitats (marine muddy sands/sandy muds, 0.1 m² grab with 1 mm sieve mesh) with corresponding metric E(X) values

Metric	% <63µm	% 63≤125µm	% 125≤250µm	% 250≤500µm	% 500≤1,000µm	% 1,000≤2,000µm	% 2,000≤4,000µm	% 4,000≤8,000µm	% ≥8,000µm	Salinity average	Salinity SD	Metric E(X) value
Taxa number												24.3
1-(AMBI/7)	50.3	21.5	20.3	6.0	1.2	0.7	0	0	0	32.4	1.2	0.65
1-λ′												0.76

Note: SD = standard deviation

The established reference conditions and estimated metric E(X) values were applied to Equation 4.5 to derive the reference condition constants (Table 4.10).

Table 4.10 Metric reference condition values for EUNIS A5.2/A5.3 (marine muddy sands/sandy muds, 0.1 m² grab with 1 mm sieve mesh) with corresponding estimated metric E(X) values and derived reference condition constants

Metric	Metric value for fully marine subtidal fine sand/mud (0.1 m² grab with 1 mm sieve mesh)		Reference condition constant
	Reference condition (expert judgement)	Metric E(X) value estimated from model	
Taxa number	78.6	24.3	3.24
1-(AMBI/7)	0.96	0.65	1.48
1-λ′	1.02	0.76	1.34

4.7 Calculation of reference condition values

Rearranging Equation 4.5, the metric E(X) values were multiplied by the reference condition constant to calculate the reference condition values:

$$E(X) \times \text{Reference condition constant} = \mathrm{X}_{\mathrm{Ref}} \qquad \text{Equation 4.6}$$

The relationship between the observed, estimated and reference condition metric values for the 0.1 m² grab and 3 × 0.01 m² core methods (1 mm sieve mesh) are presented in Figures 4.22 to 4.27. The average estimated value occurs where the observed metric value equals the estimated metric value.

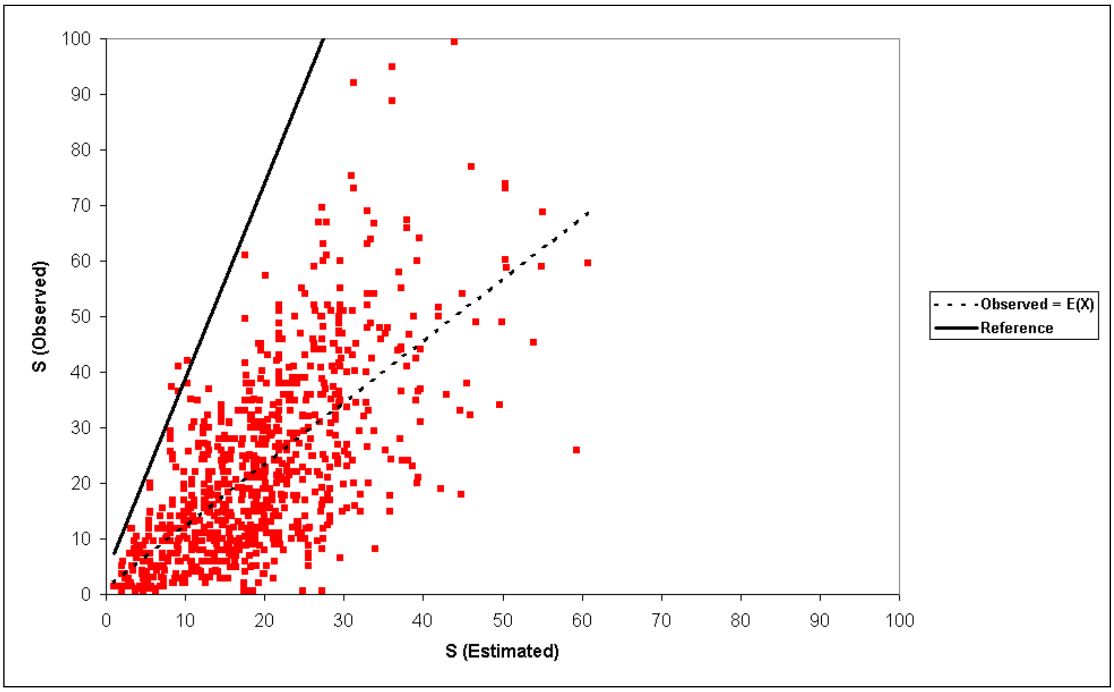


Figure 4.22 Observed versus estimated taxa number with associated reference condition value (0.1 m^2 grab, 1 mm sieve mesh)

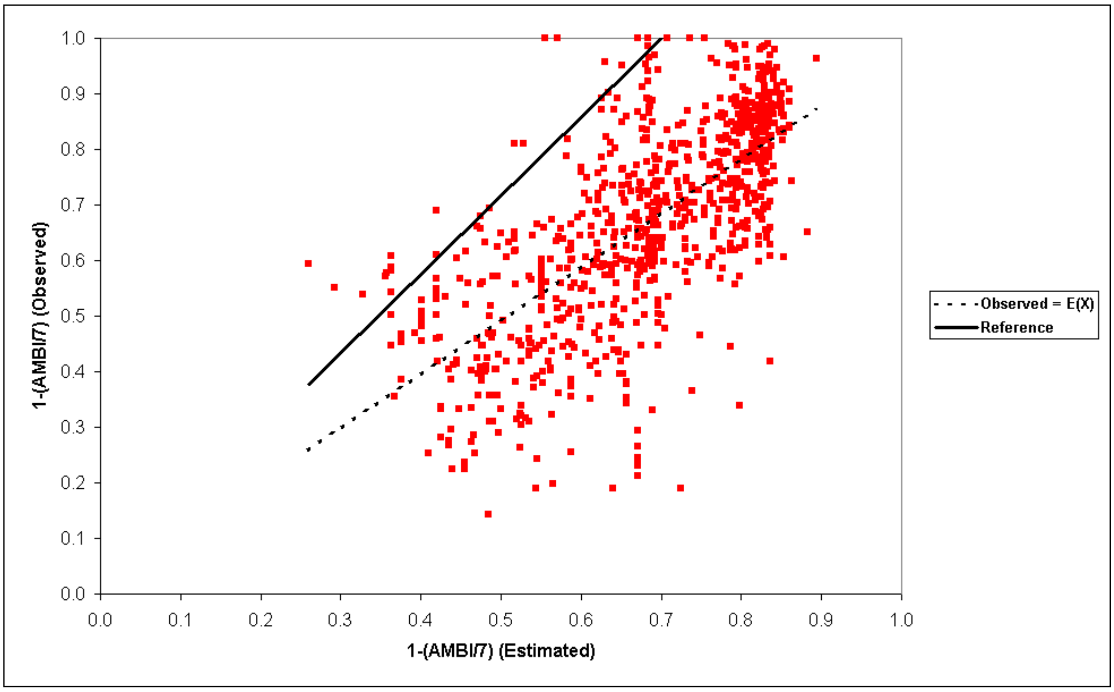


Figure 4.23 Observed versus estimated 1-(AMBI/7) values with associated reference condition value (0.1 m^2 grab, 1 mm sieve mesh)

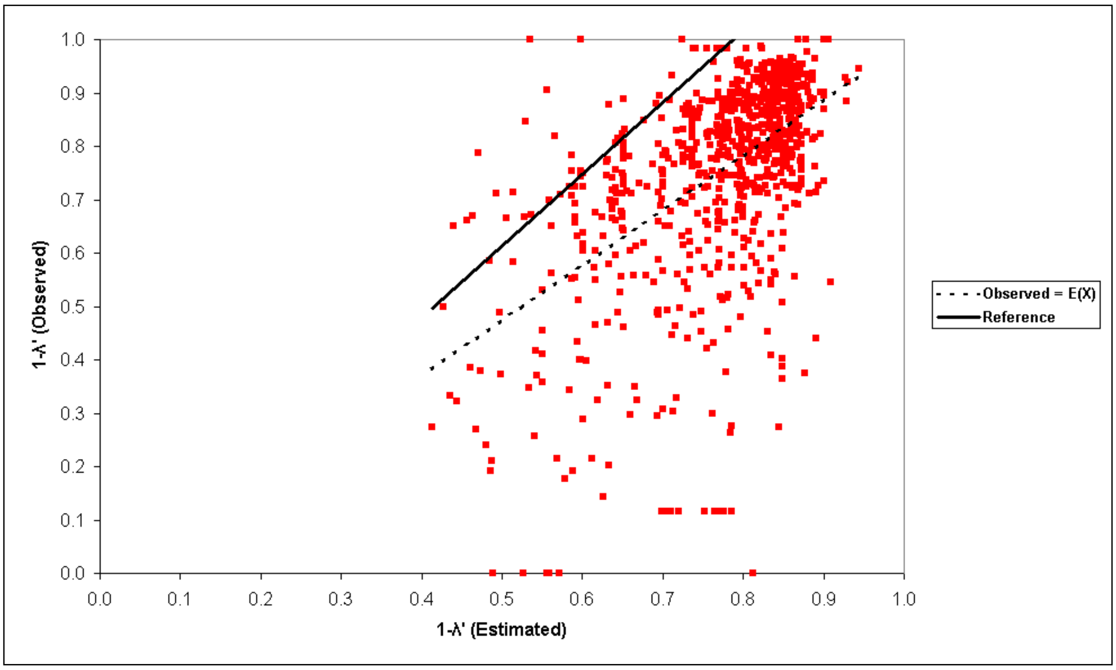


Figure 4.24 Observed versus estimated 1-λ′ values with associated reference condition value (0.1 m^2 grab, 1 mm sieve mesh)

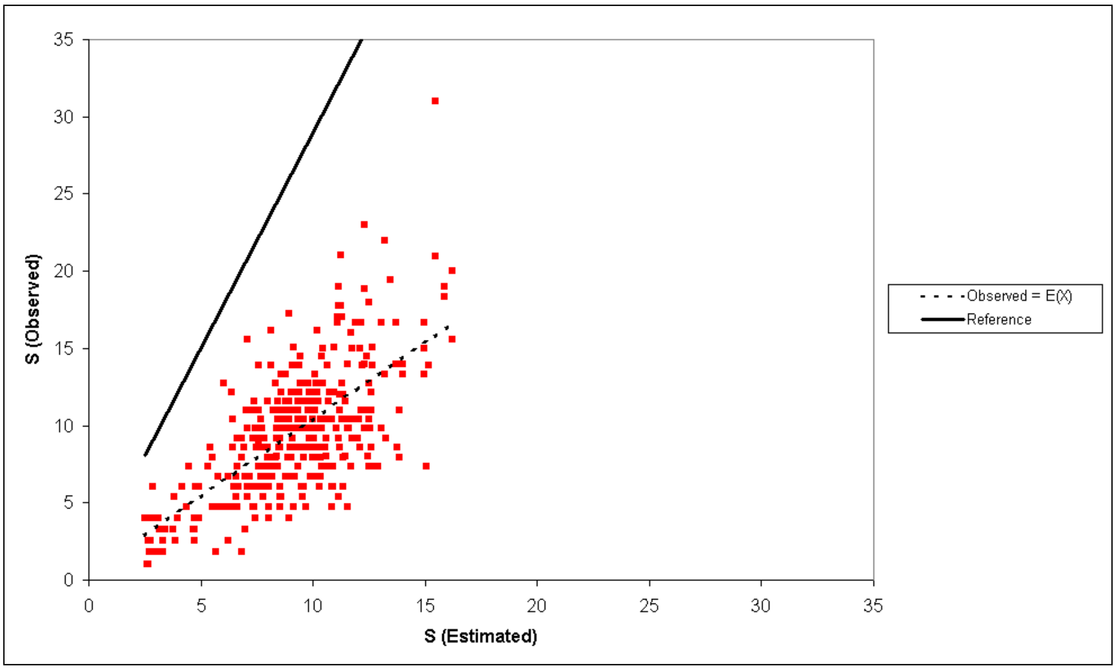


Figure 4.25 Observed versus estimated taxa number with associated reference condition value (3 × 0.01 m^2 core, 1 mm sieve mesh)

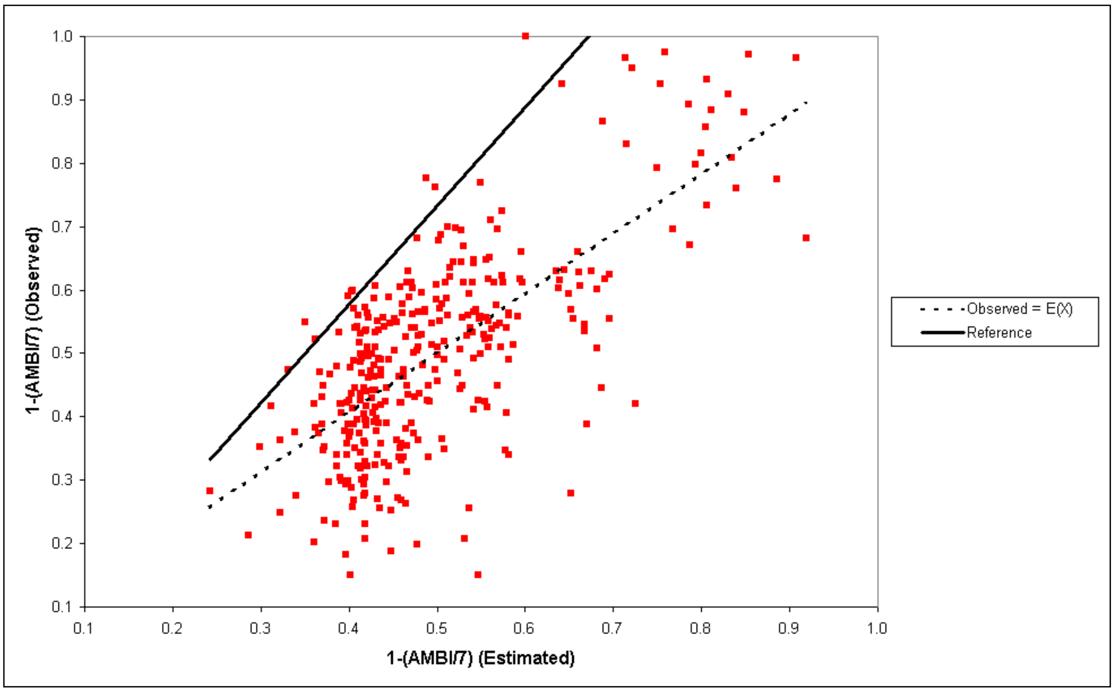


Figure 4.26 Observed versus estimated 1-(AMBI/7) values with associated reference condition value (3 × 0.01 m^2 core, 1 mm sieve mesh)

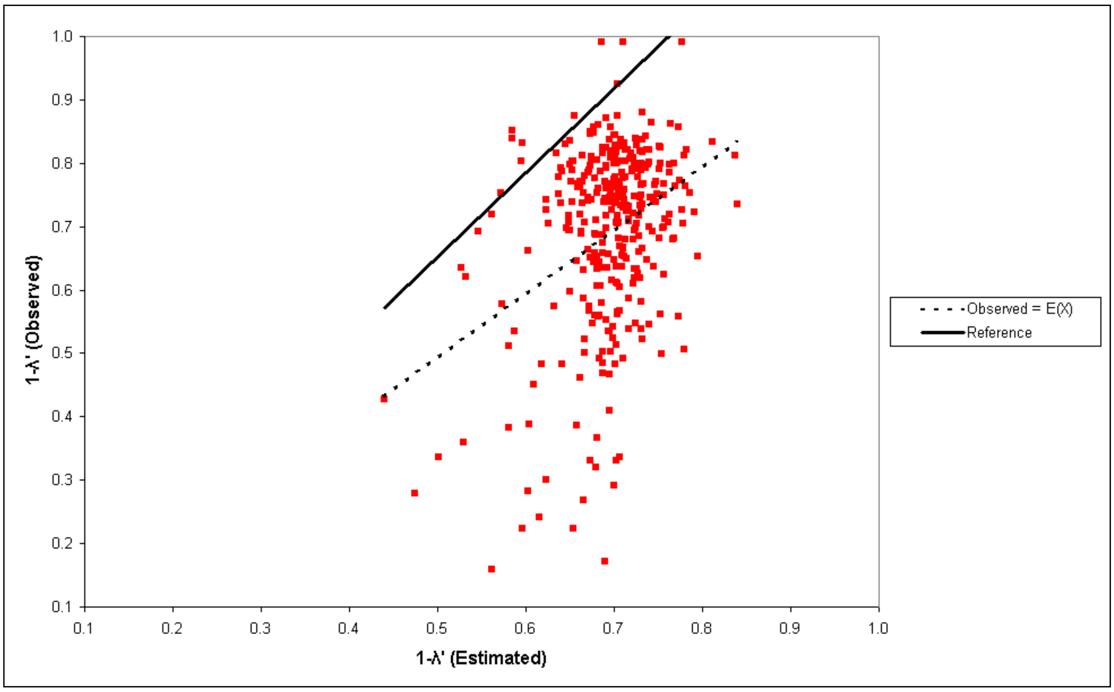


Figure 4.27 Observed versus estimated 1-λ′ values with associated reference condition value (3 × 0.01 m^2 core, 1 mm sieve mesh)

4.8 Inclusion of qualitative habitat descriptions

Although the subjectivity of the reference condition value for a sample is reduced by incorporating quantitative particle size analysis (PSA) and salinity data in the reference condition models, the absolute requirement for quantitative grain size fraction and salinity data would limit the benthic infaunal survey data available for

WFD assessment. Many infaunal surveys only record qualitative sediment descriptions; the collection of qualitative rather than quantitative sediment data is often favoured to minimise costs. Other infaunal surveys record quantitative particle size fractions different to those on which the models are based. Without a mechanism for estimating reference conditions for these records, such data are unavailable for assessment with the IQI.

To be able to utilise benthic invertebrate surveys where only qualitative sediment descriptions are available, the relationship between quantitative PSA data and sediment descriptions was investigated. As each benthic infaunal sample within the Environment Agency's WFD surveillance monitoring programme (2007-2009) has an associated PSA sample, the PSA data were used to assign a qualitative sediment description to the sample in accordance with the Folk scale (Folk 1954). The average percentage contribution for each grain size fraction was calculated for each sediment description (Table 4.11), enabling an approximation of each size fraction for qualitative sediment descriptions.

Table 4.11 Sediment descriptions with associated average grain size [1,2]

Folk (1954) description	n	Grain size fraction (μm) contribution (%)								
		<63	63<125	125<250	250<500	500<1,000	1,000<2,000	2,000<4,000	4,000<8,000	>8,000
Slightly gravelly muddy sand	35	32.5	23.4	24.8	10.6	3.6	1.9	0.9	1.4	1.1
Slightly gravelly sand	42	2.4	9.1	45.5	22.8	4.7	1.4	1.8	0.9	0.3
Slightly gravelly sandy mud	58	74.2	13.4	6.2	1.3	0.5	1.1	1.8	0.9	0.6
Gravelly mud	20	66.0	10.9	7.2	3.4	1.3	2.3	4.0	3.9	1.1
Gravelly muddy sand	25	21.6	14.5	21.0	19.6	11.2	3.0	4.1	3.6	1.3
Gravelly sand	29	1.7	8.2	43.6	33.4	4.5	1.4	3.0	3.2	1.2
Mud	55	93.9	5.0	0.9	0.2	0.0	0.0	0.0	0.0	0.0
Muddy gravel	1	47.7	10.4	9.7	0.5	0.5	0.7	1.8	28.7	0.0
Muddy sand	357	31.6	26.2	30.7	9.5	1.3	0.8	0.0	0.0	0.0

Folk (1954) description	n	Grain size fraction (µm) contribution (%)								
		<63	63<125	125<250	250<500	500<1,000	1,000<2,000	2,000<4,000	4,000<8,000	>8,000
Sand	408	1.7	10.7	50.7	33.3	3.0	0.6	0.0	0.0	0.0
Sandy mud	354	69.2	16.7	9.8	2.6	1.0	0.6	0.0	0.0	0.0

Notes: [1] It is acknowledged that reference conditions based on habitats with low numbers of samples may be unreliable. Additional data are required to improve confidence in the reference condition models for these habitats.
[2] Data from 2007-2009 Environment Agency WFD surveillance monitoring programme

Data based on the conversion of qualitative descriptions to quantitative PSA values should be used only for assessment purposes and not included in establishing reference condition values. This is because qualitative sediment descriptions are prone to subjectivity, so by excluding qualitative data the objectivity of the reference condition models is maintained. Descriptions may be influenced by factors such as the sediment/habitat classification system referred to (EUNIS, Folk, MNCR and so on), the component of the sediment observed (for example, whether descriptions are based on the appearance at the sample surface versus profile of sediment) and individual interpretation.

While the reference condition values are based on environmental parameters operating over continuous scales, the output of the approach is illustrated in Figures 4.28 to 4.33, which provide approximate reference condition values for the IQI component metrics over a range of sediments (as classified according to Folk 1954) over a range of salinity. The reference condition values for the illustrations should be considered as approximate and are to be revised with additional data when available.

The latest models used to calculate reference conditions from specific environmental parameters are available from the UK WFD competent authorities.

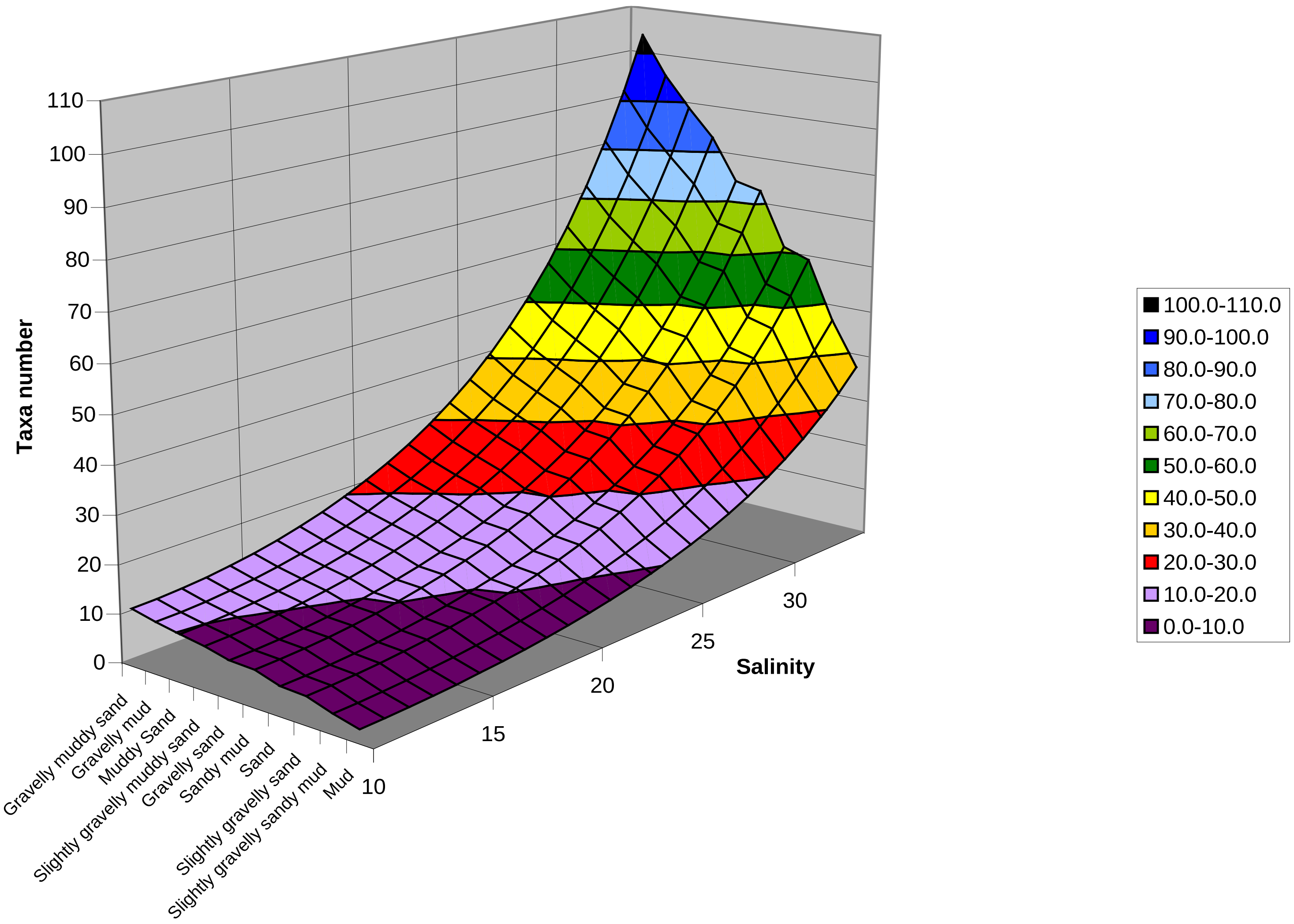


Figure 4.28 Reference conditions for taxa number (0.1 m^2 grab processed using a 1 mm sieve) for a selection of sediment types (Folk 1954) over the 10–34 salinity range, estimated using the continuous habitat reference condition approach

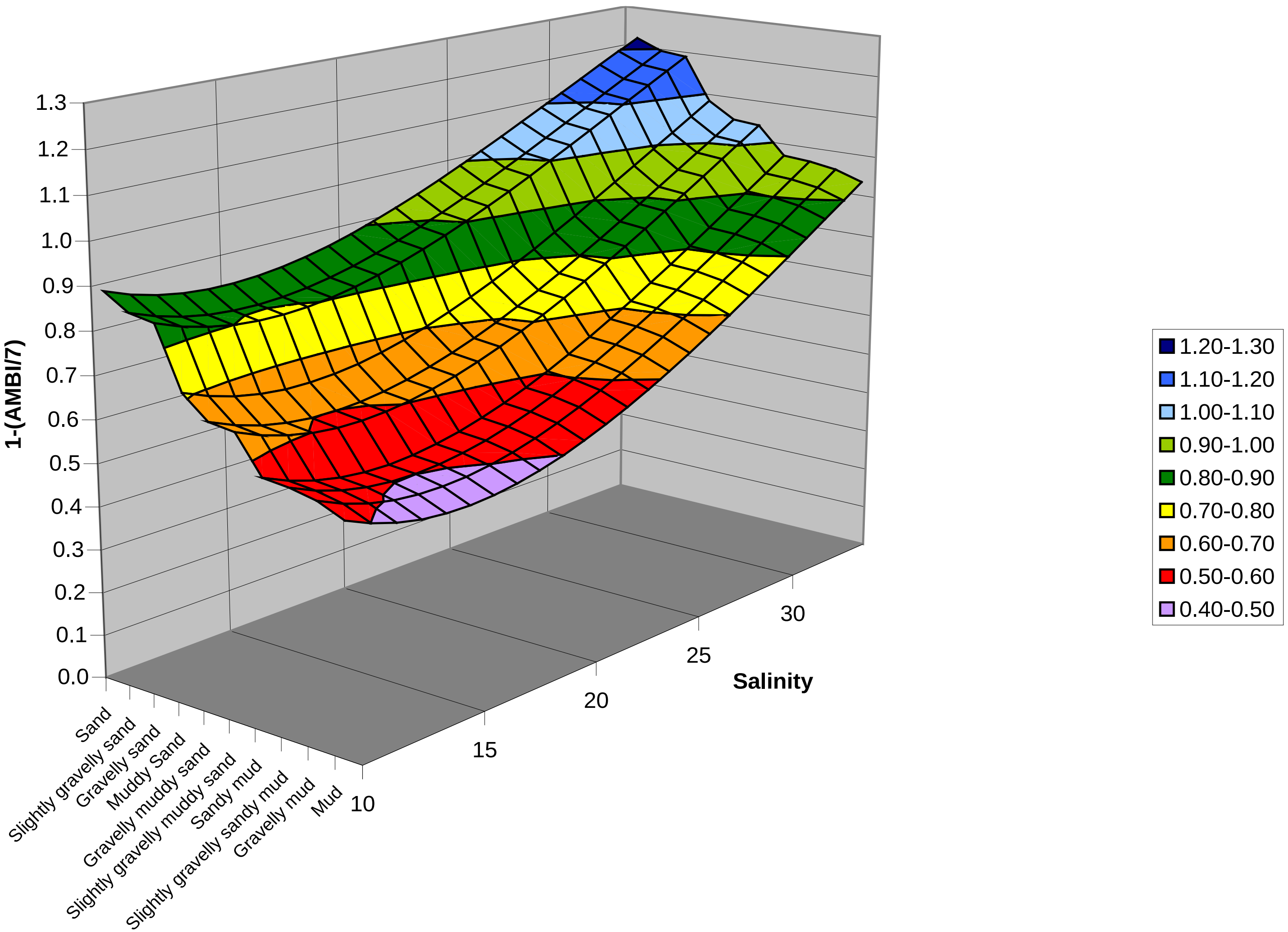


Figure 4.29 Reference conditions for 1-(AMBI/7) (0.1 m^2 grab processed using a 1 mm sieve) for a selection of sediment types (Folk 1954) over the 10–34 salinity range, estimated using the continuous habitat reference condition approach

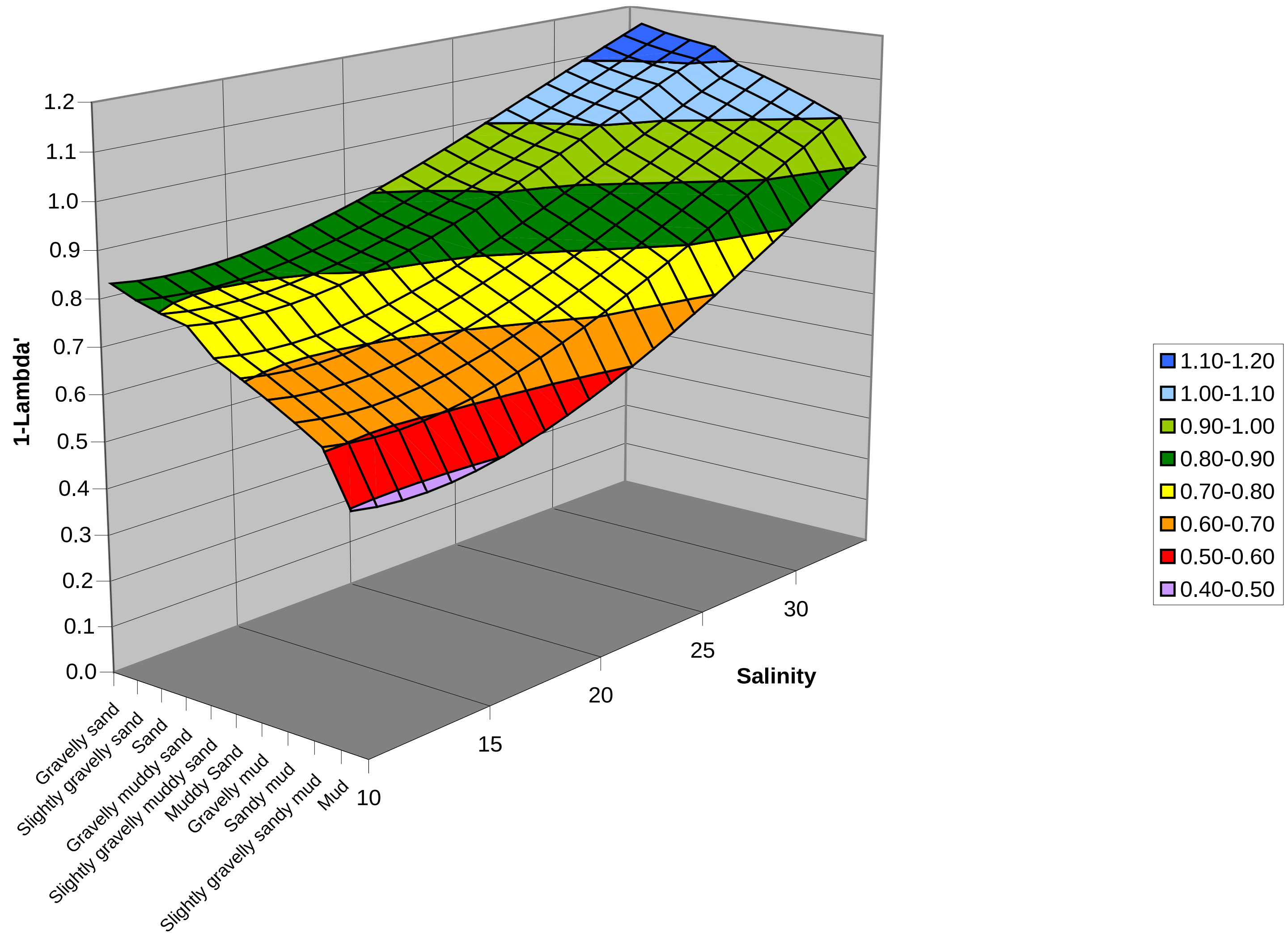


Figure 4.30 Reference conditions for 1-λ′ (0.1 m^2 grab processed using a 1 mm sieve) for a selection of sediment types (Folk 1954) over the 10–34 salinity range, estimated using the continuous habitat reference condition approach

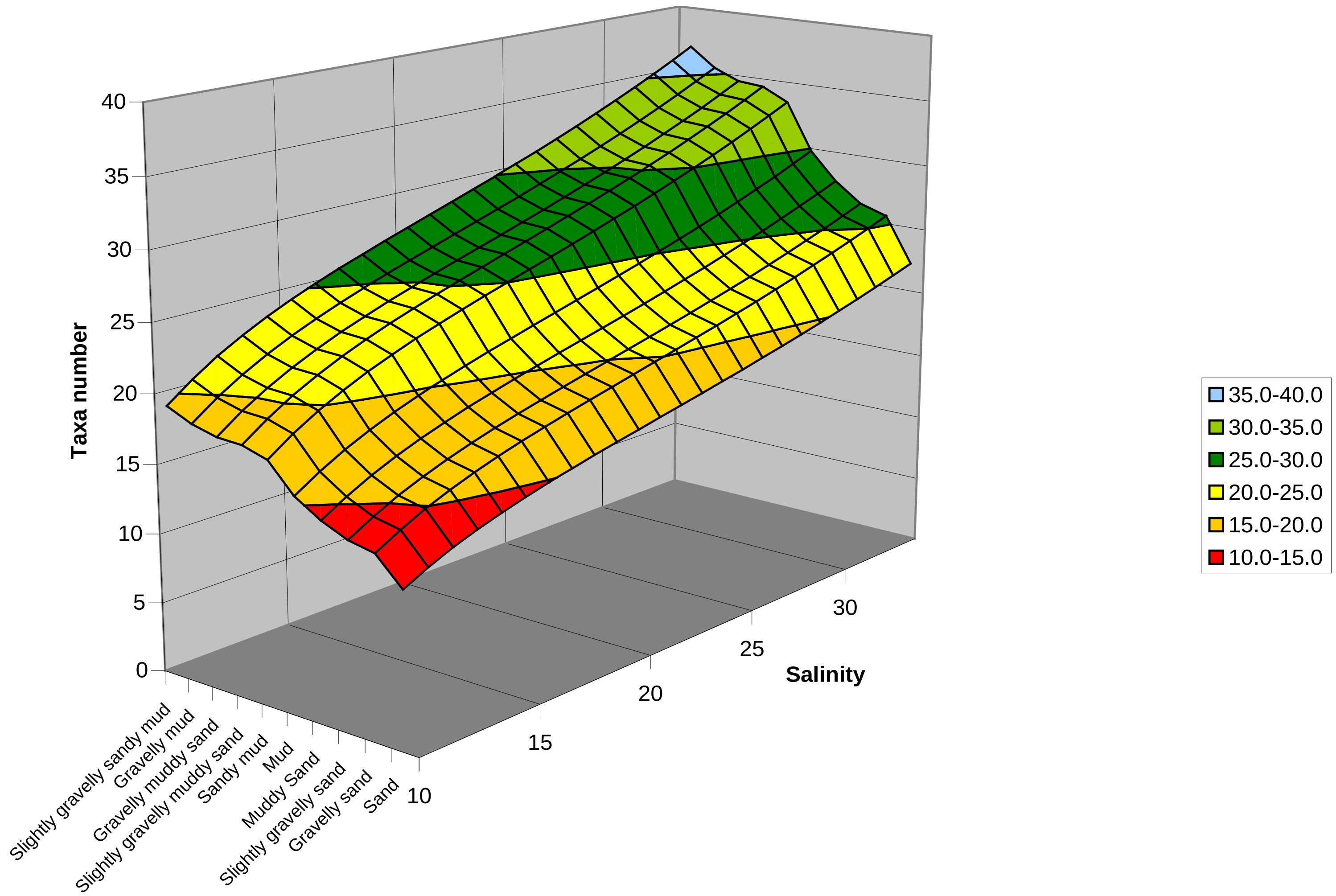


Figure 4.31 Reference conditions for taxa number (3 × 0.01 m^2 core processed using a 1 mm sieve) for a selection of sediment types (Folk 1954) over the 10–34 salinity range, estimated using the continuous habitat reference condition approach

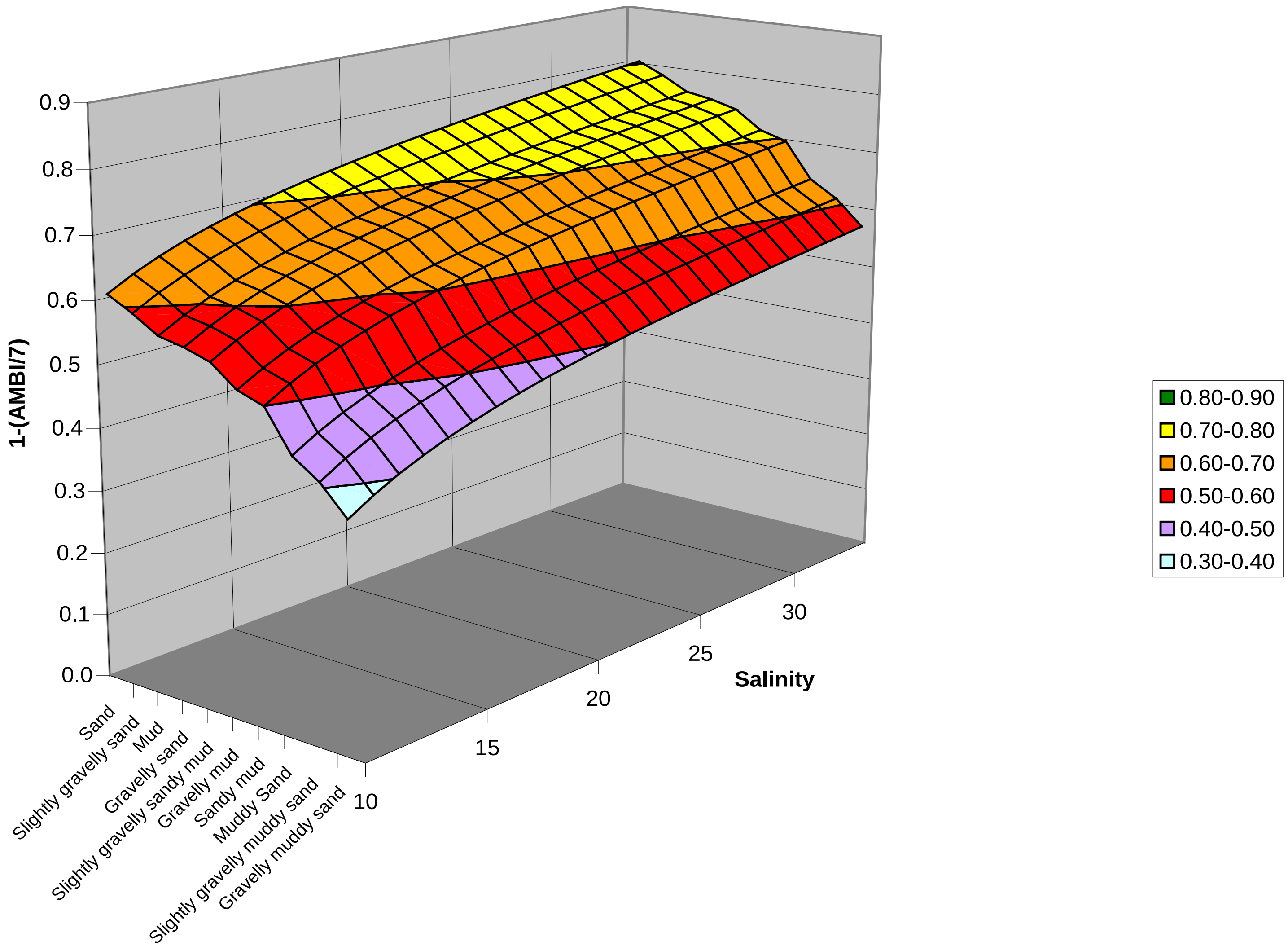


Figure 4.32 Reference conditions for 1-(AMBI/7) (3 × 0.01 m^2 core processed using a 1 mm sieve) for a selection of sediment types (Folk 1954) over the 10–34 salinity range, estimated using the continuous habitat reference condition approach

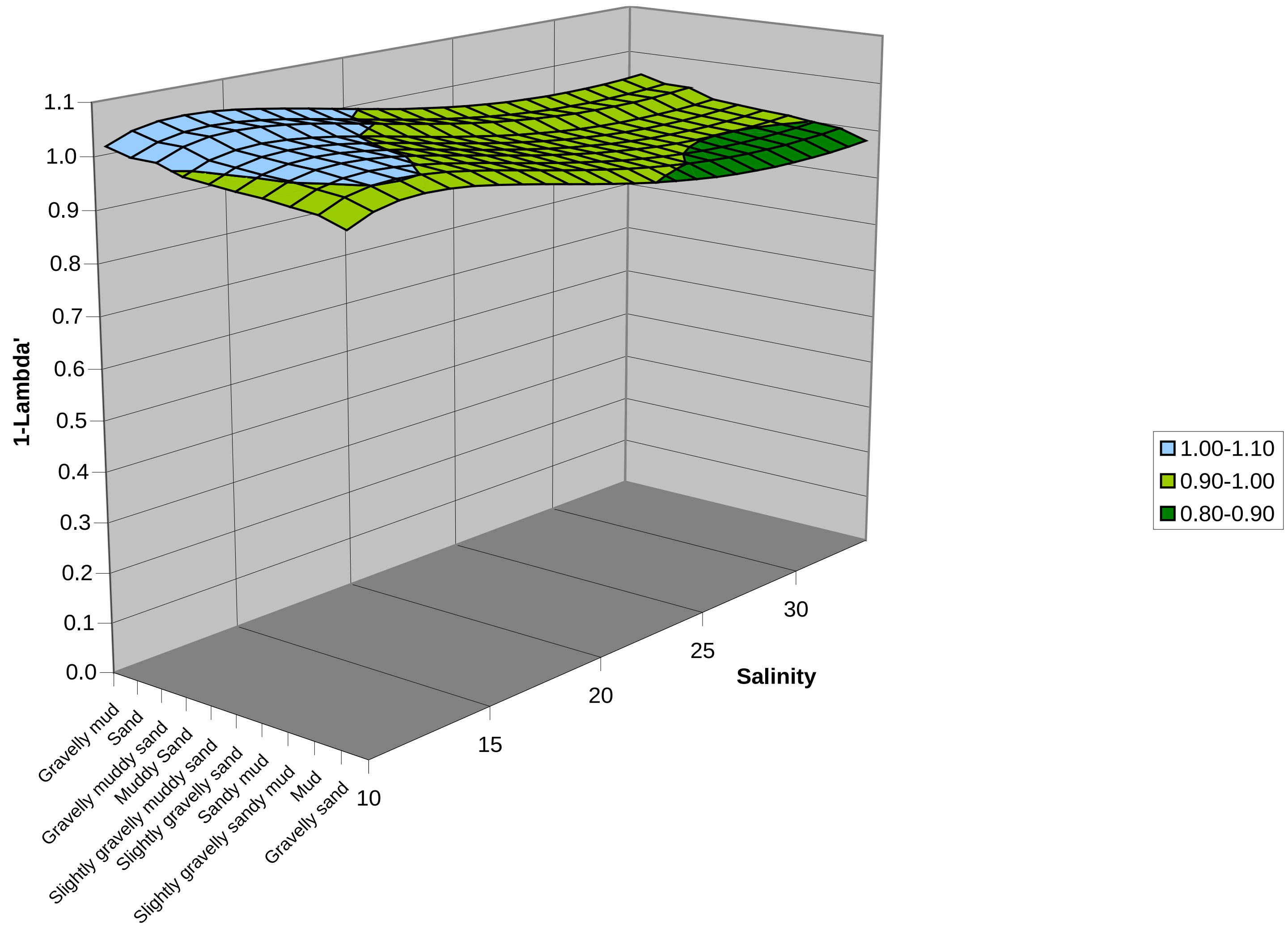


Figure 4.33 Reference conditions for 1-λ′ (3 × 0.01 m^2 core processed using a 1 mm sieve) for a selection of sediment types (Folk 1954) over the 10–34 salinity range, estimated using the continuous habitat reference condition approach

4.9 Discussion

The approach described above is based on the currently available biological and supporting data from UK national monitoring programmes. Additional ecological factors may also influence the IQI metrics which have not so far been addressed. They must be acknowledged as being potential sources of error and bias in the devised reference conditions (consequentially translating into bias and error in the associated IQI results) and will be reviewed as additional data become available.

4.9.1 Correlation between anthropogenic and natural conditions

Although use of all available data in modelling reference conditions reduces the subjectivity when adapting reference conditions across a range of natural environmental conditions, it assumes that there is no correlation between anthropogenic pressure and the physicochemical parameters on which the models are based. This is not expected to be the case as there are anticipated links between the extent of anthropogenic pressure and certain physicochemical parameters. Examples include the correlation between pressure and salinity as a result of relatively high levels of development and activity within transitional waters, and the correlation between certain contaminants and silt/clay content as a result of contaminant adsorption being elevated for particles with high surface area:volume ratios.

To test the extent to which this may be the case (thus potentially providing an alternative basis to substitute this assumption with, for example, the development of pressure correction factors) would require the compilation of data on the pressures relevant to the IQI metrics to be related to the physicochemical parameters used in the analysis. Without such information, and therefore applying the above assumption, the models are unable to differentiate between the effects of natural and anthropogenic stress on the metrics in instances where the anthropogenic pressure and physicochemistry correlate.

4.9.2 Effects of biogeography

Marine taxa have an optimum biogeography where they are most able to compete against other taxa for resources and have the most suitable conditions for reproduction, typically linked to physicochemical (temperature, light, geographical isolation and so on) and ecological (resource competition, predation and so on) conditions. The tolerance of taxa to anthropogenic disturbance is likely to depend on their proximity to this optimum biogeography, with the likelihood being that taxa are more sensitive at their latitudinal distribution limit (Grall 2007).

This may have implications to AMBI sensitivity scores (predominantly based on taxa distributed throughout the North East Atlantic maritime ecoregion – the AMBI taxon list is available at http://ambi.azti.es). For example, a taxon allocated to AMBI ecological group III (taxa tolerant to excess organic matter enrichment) may be relatively disturbance tolerant at its biogeographical optimum, but may become increasingly sensitive to disturbance as it approaches the limits of its biogeographical distribution. Under such circumstances allocating an AMBI ecological group of I or II may be more appropriate. This has potential implications when using taxon sensitivity scores within assessments over large spatial scales, for example, across the Member States of the North East Atlantic Intercalibration Group.

While it may be beneficial for the effect of shifting taxon sensitivity to be factored into the AMBI assessment, the actual implications of biogeography on AMBI are lessened by acknowledging that for any single location across a geographical gradient, AMBI values may be considered comparable as long as there is an approximately constant balance of taxa in their biogeographical optimum and sub-optimum. This principle may be less applicable at locations at the extremes of the geographical area from where the taxa within the AMBI list are distributed. In such areas, it may be that the majority of taxa within the AMBI taxon list are in their biogeographical sub-optimum.

4.9.3 Pressure (pollution) induced tolerance

Under severe pollution stress, the ability to undergo short-term genetic selection enables certain taxa to dominate (typically r-strategists, expressed within AMBI as ecological group IV and V taxa). At the population level, localised chronic exposure of macrobenthic taxa to certain anthropogenic pressures has resulted in elevated levels of resistance in localised populations such as the increased resistance of nematodes to copper from populations with historical exposure to severe contamination (see Millward and Grant 1995). This has the potential to cause inconsistencies in using biological community measures to monitor and assess the extent of anthropogenic pressure; the exposure of two structurally identical communities (one adapted to historic chronic exposure to a given pressure, the other devoid of adaptations) to a fixed level of anthropogenic pressure may provide differing ecological assessment results. While the extent to which this aspect affects the metrics within the IQI is expected to be negligible, it should not be disregarded fully without further study.

This scenario is permitted under the normative definitions of the WFD where ecological status is defined as the extent departure from reference conditions of the biological community. An assemblage of a given structure would not reflect the same extent of anthropogenic pressure in the case where differing degrees of pressure result in the same extent of departure.

4.9.4 Seasonality

Macrobenthic assemblages are influenced by seasonality (changes in temperature, food availability, riverine flow rates and so on). Such influences may induce elevated abundance or presence of certain taxa due to the recruitment of juveniles. Additional seasonally variable influences to macrobenthic assemblages also exist such as the removal of certain taxa by predation by avifauna or piscifauna.

Further work is required establish whether such changes to macrobenthic assemblages are reflected by the results of IQI assessments. Currently, this potential effect is reduced through survey design. WFD benthic infaunal monitoring is restricted to a fixed time window from February to May (inclusive), minimising the potential bias caused to the results as a result of seasonality. While the existing study into the effect of juveniles on the IQI (Chapter 2) has indicated that they cause no significant effect on EQRs, this is restricted to data from the CSEMP programme, sampled between February and May (inclusive) where the main recruitment phase for many macrobenthic invertebrate taxa is avoided. The influence of juveniles to EQRs within data sampled outside the February to May window, where juvenile recruitment may be more influential, is not yet known.

4.9.5 Inclusion of additional data

The effectiveness of the regression models for estimating metric values is increased by the number of data points. For hydromorphological and physicochemical conditions where little data exist (particularly under highly variable conditions), the precision of the models is reduced. This can be observed by relating taxa number to salinity for Environment Agency 2007-2010 WFD data (Figure 4.34).

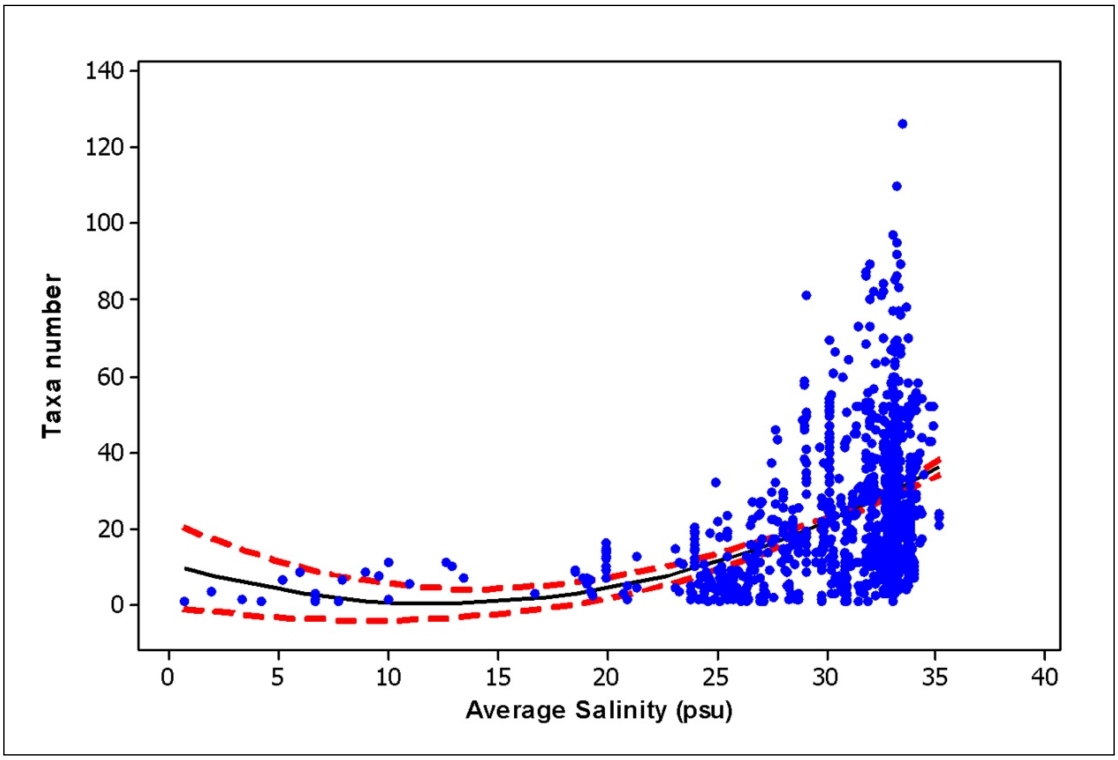


Figure 4.34 Taxon number (truncated) versus average salinity (Environment Agency 2007-2010 WFD data) with fitted quadratic regression with upper and lower 95% confidence intervals

The 95% upper and lower confidence intervals (CIs) fit tightly to the mean for the marine data (salinity = ~33) where the standard error is lowest (high number of data values for a given standard deviation). At salinity <20, most data fall below the upper 95% CI, that is, confidence that the maximum values observed are greater than the average fitted through the data (comparable with the metric e_s) is <95%. As the reference condition values are a function of a fitted regression through the data points, at low salinity (that is, approximately <20) the reference condition values calculated from the models should be treated with low confidence until sufficient data exist for such conditions. This situation applies to all metrics and all physicochemical parameters.

WFD Annex II section 1.3(vi) states that where sufficient levels of confidence cannot be attained for reference conditions, a classification tool may be unsuitable for assessment:

> 'where it is not possible to establish reliable type-specific reference conditions for a quality element in a surface water body type due to high degrees of natural variability in that element … then that element may be excluded from the assessment of ecological status for that surface water type'.

There are plans to update the reference models with additional data as they become available in order to improve the reliability of the models so as to enable the IQI to be applied to as broad a range of environmental conditions as possible. This is important to ensure that as broad a range of environmental conditions as possible can be proportionately represented when assessing the overall ecological quality of the benthic assemblage within a water body.

The variability of the modelled IQI metric values (and corresponding confidence that the values estimated from the models fall between an upper and lower confidence interval) will have implications to the overall risk of misclassification (Chapter 6). The variability of reference conditions is not currently incorporated within the risk of misclassification, but are likely to have implications to the derived EQR values, particularly under extremes of environmental conditions where little data are available as illustrated for low salinity environments (Figure 4.34). Consideration must be given to the certainty of the reference condition when suggesting management action.

4.9.6 Additional hydromorphological and physicochemical factors

The current reference condition models are based on a limited set of environmental parameters and will undoubtedly exclude other parameters influential to changes in metric reference condition values as a result of natural systematic bias. The models should only be considered effective in differentiating reference condition values in accordance with changes in those environmental parameters included.

Using the reference condition models described in this report, the extent to which variability (R^2) can be estimated from the environmental parameters selected ranges from ~0.51 (taxa number for 3×0.01 m^2 intertidal core data) to ~0.08 ($1-\lambda'$ for 3×0.01 m^2 intertidal core data). Where models estimate a relatively low degree of metric variability, this indicates that one or more of the following factors may apply.

- The physicochemical parameters selected are insufficient and the use of additional parameters should be explored.
- The metric is less influenced by environmental factors in contrast to anthropogenic factors.
- The metric is more susceptible to random error and sampling and measurement error.

Certain hydromorphological and physicochemical factors that are considered important drivers of macrobenthic assemblages have not currently been included in developing the reference condition models listed in Section 4.6. This has predominantly been due to data availability. The parameters listed are being reviewed to ensure their suitability in revisions for the reference condition models, with the view to adding further natural environmental parameters that may account for significant additional variability in the metrics, potentially improving the ability of the models to estimate values at reference conditions. Such factors to be considered are outlined below.

Inclusion of standardised depth data

Depth is an important driver in coastal systems and has been shown to correlate with the metrics within the IQI, such as a reduction in species number with depth (Gray et al. 1997) and the influence on diversity values with depth (Simboura and Zenetos

2002). It is expected that metric reference values are also likely to follow this correlation. However, the samples within the current UK WFD dataset are relatively shallow when compared with these studies (depth >30m in <0.5% of records) and therefore the relationship may not be as pronounced.

Although a common supporting parameter to macrobenthic data and routinely recorded, depth is often recorded as absolute depth below vessel and not depth relative to chart datum. Existing bathymetric data may be used as an alternative, but are often based on historic records (for example, Admiralty chart data) with data points often sparsely distributed, greatly reducing the reliability of such records in estimating depth to chart datum.

Inclusion of PSA statistics

Particle size analysis statistics (for example, mean grain size and sorting coefficient) have been highlighted as important drivers in macrobenthic assemblage structure, for example, faunal diversity has been shown to increase with the diversity of particle sizes (Etter and Grassle1992), reflecting changes in habitat complexity/niche diversity. Mean, Kurtosis, skew and sorting coefficients are being investigated for inclusion in future reference condition model revisions. These were not initially incorporated as certain inconsistencies were identified in their calculation between different data sources. The quality assurance of the PSA data has been addressed by the NMBAQC (2009).

Improved salinity data

The salinity data within the reference condition models were based on spot samples. However, such data may provide an incomplete picture of the characteristics of the salinity at a site, that is, the frequency of sampling and methodology may be insufficient to capture those salinity characteristics at a site which influence the IQI metric values. For example, high stress may be induced by naturally occurring short periods of low salinity which may fall entirely between consecutive sampling events. This scenario exists within the tidal Bann transitional water body of Northern Ireland where the water body infrequently experiences several days of naturally occurring low salinity, which is sufficient in altering the benthic assemblages to a state comparable with that exposed to anthropogenic impacts (Mackie T, personal communication, 2009). Where neither sufficient salinity data nor sufficient means of factoring such natural events into the reference condition models exist, the ecological quality of such sites may be underestimated as a consequence.

4.9.7 Adaptation for subsampling methods

Certain subsampling methods can influence the IQI metric values, such as by lowering of taxa number in cases where rarer taxa are missed as a result of the process. As the AMBI index is based on proportions of sensitivity groups, it is anticipated that the effect of subsampling to AMBI values will be negligible on the basis that the proportion of AMBI ecological groups should be comparable between the full sample and subsample populations. Subsampling also has the potential to influence Simpson’s evenness whereby rare taxa are excluded from the analysis, therefore increasing the spread of the assemblage abundance over fewer taxa and thus reducing the overall evenness. Such influences would therefore also automatically apply to expected values under reference conditions.

The current reference condition values can be applied only to fully sorted and identified invertebrate data. To enable the assessment of subsampled data by the IQI, the effects of subsampling on each metric would need to be quantified and the reference condition values adapted accordingly along with the provision of appropriate protocols and guidance to enable compliant subsampling to be undertaken. The extent to which subsampling methods affect the IQI metric values has not been investigated.

4.9.8 Saline lagoons

Both coastal and transitional saline lagoon water body types are recognised with the WFD (types CW10 and TW6 respectively). They constitute important conservation features and are categorised as high priority habitats under the Habitats Directive. Saline lagoons differ from other transitional and coastal water body types as they are partially separated from the sea, resulting in differences in their physicochemical regime and associated benthic fauna. It is anticipated that the IQI could be used for the assessment of the benthic invertebrate community in saline lagoons once suitable reference conditions have been established.

4.9.9 Conversion factor for MNCR/EUNIS habitats

The MNCR (and associated Marine Habitat Classification for Britain and Ireland) and EUNIS habitat classification systems are widely adopted throughout the UK and Ireland, and Europe respectively. They are of particular significance in habitat definition for conservation purposes (Habitats Directive) and are frequently recorded to support macrobenthic data. To allow the reference condition values to be more reliably estimated for macrobenthic data supported by MNCR/EUNIS habitat classification data, appropriate values for the habitats defined under these systems require estimating, as undertaken for the Folk sediment classification system (0). However, as the MNCR and EUNIS classification systems use terms or parameters to define habitats that are difficult to express quantitatively (such as the division of sediments into terrigenous muds), the reference models may be unable to appropriately differentiate reference conditions accordingly.

5 IQI class boundary setting

The ecological health of the biological quality elements are categorised as ecological status classes under the WFD. This chapter describes how the WFD normative definitions for each ecological status class were interpreted into numerical class status boundaries. It also outlines how boundaries have been modified to ensure comparability of boundaries with other Member States across the North East Atlantic intercalibration process.

5.1 Introduction

The WFD requires the classification of ecological quality elements into five distinct ecological status classes: high, good, moderate, poor and bad. The normative definitions (Table 5.1) provide the specific aspects of each quality element to be considered and a qualitative description of those parameters in the different status classes.

Table 5.1 Normative definitions (as outlined in WFD Annex V section 1.2) for the classification of the benthic invertebrate quality element into five ecological status classes

Ecological status	Normative definition
High	• The level of diversity and abundance of invertebrate taxa is within the range normally associated with undisturbed conditions. • All the disturbance-sensitive taxa associated with undisturbed conditions are present.
Good	• The level of diversity and abundance of invertebrate taxa is slightly outside the range associated with the type-specific conditions. • Most of the sensitive taxa of the type-specific communities are present.
Moderate	• The level of diversity and abundance of invertebrate taxa is moderately outside the range associated with the type-specific conditions. • Taxa indicative of pollution are present. • Many of the sensitive taxa of the type-specific communities are absent.
Poor	• Major alterations to the values of the biological quality elements for the surface water body type. • Relevant biological communities deviate substantially from those normally associated with the surface water body type under undisturbed conditions.
Bad	• Severe alterations to the values of the biological quality

Ecological status	Normative definition
	elements for the surface water body type. • Large portions of the relevant biological communities normally associated with the surface water body type under undisturbed conditions are absent.

The results from the monitoring of the biological quality elements are compared to reference conditions and expressed as an EQR (Figure 5.1). The level of deviation from reference condition defines the ecological class.

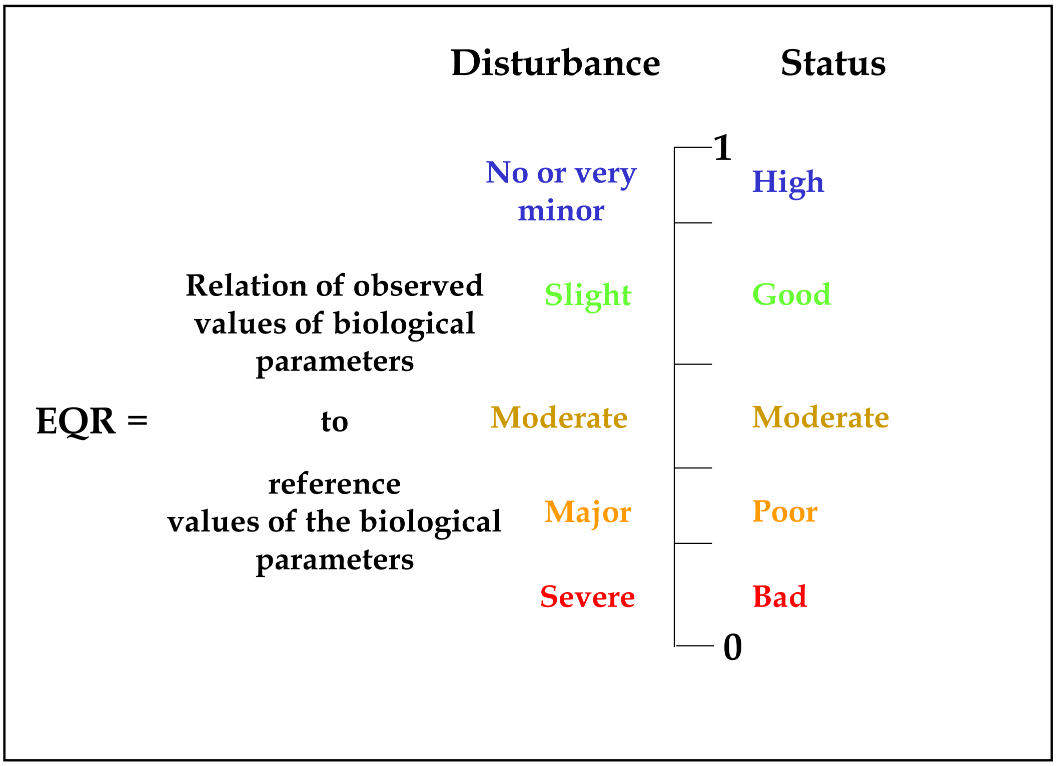


Figure 5.1 Suggested EQR according to WFD Annex V section 1.4.1

Note: The size of each status band differs because the boundaries between classes must align with the normative definitions, not a simple percentage (COAST 2003).

Class boundaries represent the extent of departure from reference conditions resulting from anthropogenic pressure. The boundaries between each of the status classes have to be defined quantitatively in terms of the criteria outlined in the normative definitions (Table 5.1). The appropriate setting of the good–high and moderate–good boundaries is key within WFD assessments as classification of a water body as moderate ecological status (or worse), or the deterioration of ecological status from high to good, drives management action.

To ensure consistency in classification across Member States, the setting of the moderate/good and good/high boundaries have been aligned through work by the NEAGIG as specified in WFD Annex V (Table 1.4.1(iii)).

5.2 Initial boundary setting (UK and RoI)

Prior to the Phase I NEAGIG Intercalibration, national boundaries were proposed for the UK and RoI. In order to develop these initial boundaries, the normative definitions of the WFD were first expanded by MBITT (Appendix B; see also Prior et al. 2004). The expanded normative definitions were derived to provide specific requirements for a range of criteria for the measurement of the ecological health of fully marine, sublittoral soft sediments (EUNIS A5.2 and A5.3 habitats).

The interim technical report explored preliminary approaches to boundary setting criteria (Prior et al. 2004). It highlighted the fact that the water body type specific reference conditions for benthic invertebrate assemblages, and therefore the associated status boundaries, must take into account the substratum and salinity regime that the assemblage inhabits (see Chapter 4).

Initial approaches explored the setting of distinct boundaries according to discrete habitat types. However, by developing the IQI so that a given EQR represents a defined proportional departure from reference condition regardless of habitat and sample methodology, the class boundaries are kept constant. This enables the same class boundary values to apply to all IQI assessments regardless of water body or habitat type.

5.2.1 Identifying boundary setting criteria

The Garroch Head sewage sludge disposal ground gradient dataset (see Chapter 3) was used to develop the initial IQI class status boundaries. The IQI combines the metrics AMBI, taxa number (S) and Simpson's evenness ($1-\lambda'$) (Chapter 3).

The underlying mechanics behind the calculation of AMBI differs to that of taxa number and Simpson's evenness. In terms of changes in the macrobenthic assemblage structure, the behaviour of taxa number and Simpson's evenness is essentially continuous; there are no distinct observable changes in assemblage composition between one value and the next to indicate that a change in ecological status has occurred. For example, a difference in taxa number of 30 versus 40 could not be related to the expanded normative definitions to objectively justify that one sample was representative of good status and the other of high status. It was therefore considered inappropriate to use taxa number and/or Simpson's evenness as the basis for setting boundaries along the pressure gradient, as the point of transition from one status to the next would be open to a high level of subjectivity.

Using AMBI represents a less subjective approach. The AMBI follows the ecological principles defined by Pearson and Rosenberg (1978) over an environmental stress (organic enrichment) gradient (Figure 5.2) and follows the model developed by Grall and Glémarec (1997). The AMBI attempts to represent the overall sensitivity of a benthic assemblage by representing a weighted average of the proportions of taxa from five different ecological sensitivity groups:

- EGI = disturbance sensitive taxa
- EGII = disturbance indifferent taxa
- EGIII = disturbance tolerant taxa

- EGIV = second order opportunists
- EGV = first order opportunists

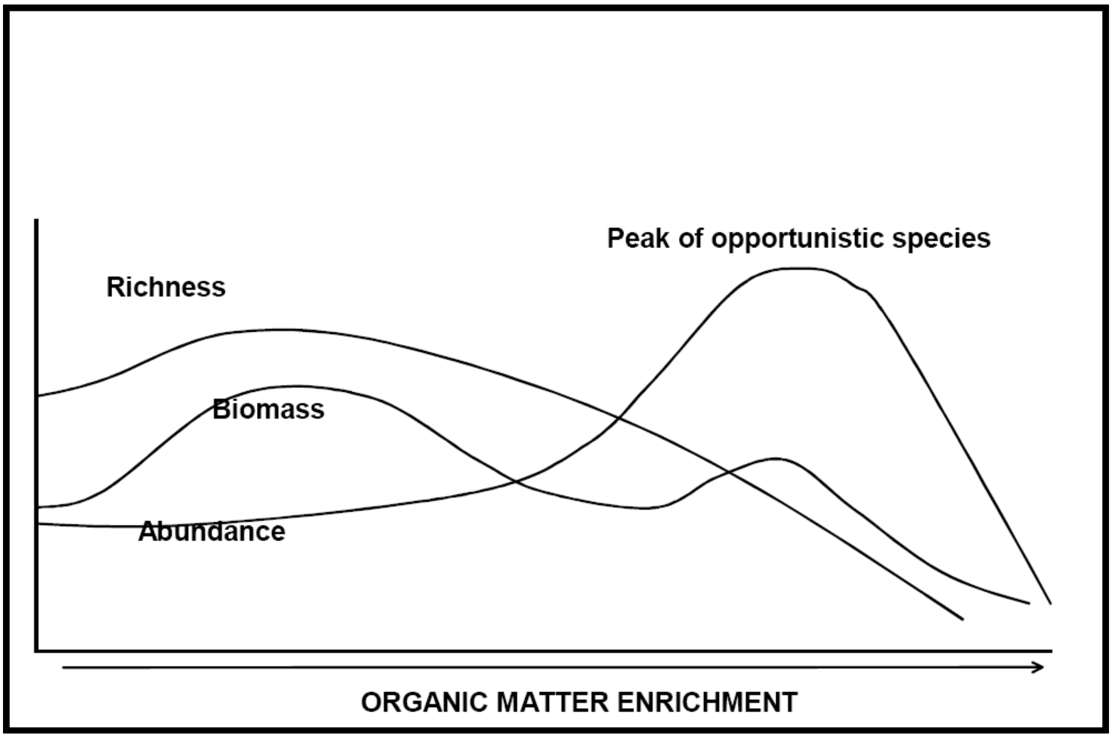


Figure 5.2 Model developed by Pearson and Rosenberg (1978) to show changes in species richness, biomass and abundance in response to an environmental stress (organic enrichment) gradient

Plotting the expected distributions of the different AMBI ecological groups over an organic pressure gradient enables the parallels between the Pearson and Rosenberg model to be observed (Figure 5.3). Borja et al. (2003) proposed corresponding WFD status classes in addition to describing the changing distribution of each ecological group. It is apparent that the observed patterns in AMBI are based on principles that can be differentiated into discrete sections over a pressure gradient compared with taxa number and Simpson's evenness.

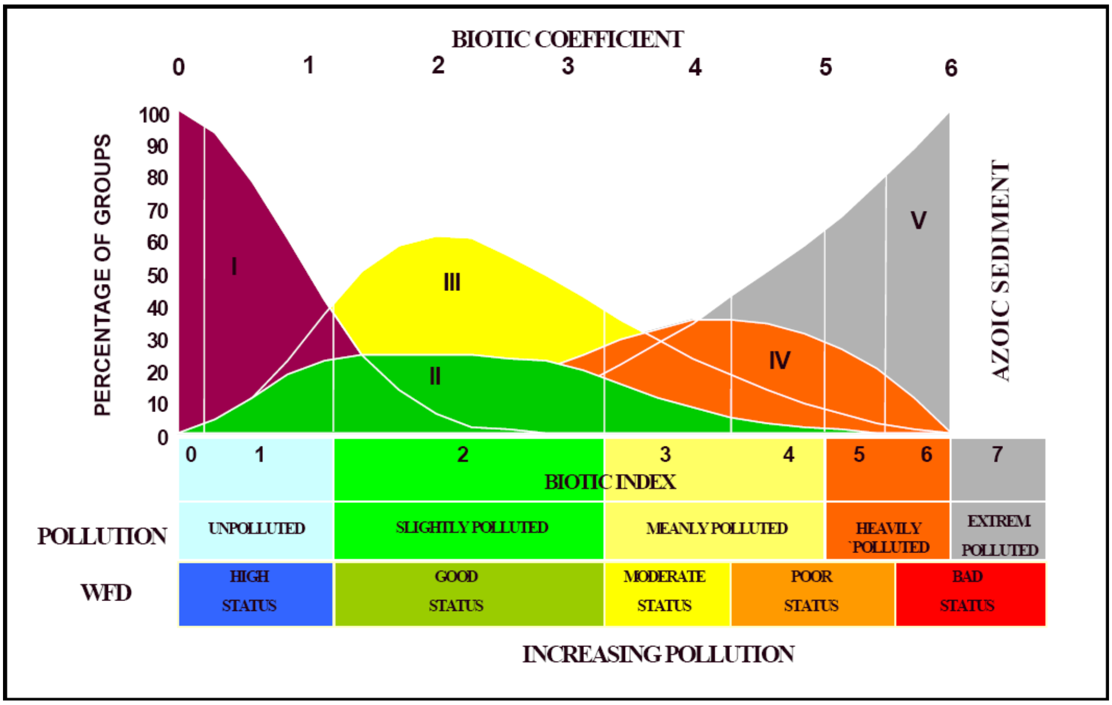


Figure 5.3 Theoretical model modified from Hily (1984), Hily et al. (1986) and Majeed (1987), which provides the ordination of soft-bottom macrofauna species into five ecological groups (EGI–EGV) according to their sensitivity to an increasing pollution gradient (from Borja et al. 2000) including the behaviour of the AMBI biotic coefficient with suggested class boundaries for WFD (Borja et al. 2003)

While AMBI sensitivity values are assigned to taxa in the form of discrete sensitivity groups (EGI–EGV), in reality the sensitivities of taxa within an assemblage are more likely to operate over a continuum. As a result, the shifts in ecological state displayed by the AMBI may also be considered as arbitrary divisions of the overall sensitivity scale. This is an issue that often applies to systems where ecological parameters that typically operate over continuous gradients are classified into distinct arbitrary, and potentially subjective, groups (see Section 4.5). Although the proposed divisions based on AMBI can be considered arbitrary to an extent, the values do provide an ecological basis with some degree of objectivity for identifying distinct boundaries over a continuous pressure gradient. As such, the boundaries can be related back to the expanded normative definitions with a relative degree of objectivity. As described by Prior et al. (2004):

> 'the advantage of the AMBI in defining ecological status boundaries is that the proportions of taxa at each stage of the organic gradient can be related to the normative definitions of all or most sensitive taxa present with respect to high and good status, respectively, and many sensitive taxa absent with regard to moderate status'.

The expected proportions of the AMBI groups for each ecological status for EUNIS A5.3 habitats (on which the status boundaries were initially established) are described in the expanded normative definitions as shown in Table 5.2.

Table 5.2 Expected AMBI ecological group abundance composition for each status class for EUNIS A5.3 habitats as specified in the expanded normative definitions

Class status	Abundance composition				
	EGI (Sensitive)	**EGII (Indifferent)**	**EGIII (Tolerant)**	**EGIV (Opportunist)**	**EGV (Indicator)**
High	Dominant	Absent or sub-dominant	Absent or sub-dominant	Absent or negligible	Absent or negligible
Good	High sub-dominant to absent	Low sub-dominant	Dominant	Negligible or low to equi-abundant with EGII taxa	Negligible or low to equi-abundant with EGII taxa
Moderate	Negligible or absent	Low sub-dominant	Co-dominant	Co-dominant	Co-dominant
Poor	Negligible or absent	Negligible or absent	Sub-dominant	Co-dominant	Co-dominant
Bad	Absent	Absent	Absent	Sub-dominant	Dominant

5.2.2 Setting EQR class boundaries

During the initial process for setting UK and RoI boundaries prior to intercalibration, class status boundaries were set for $IQI_{v.II}$. In summary, the IQI status boundaries were adapted, tested and set using a stepwise approach in which arbitrarily placed equidistant boundaries were set along the EQR scale and then adapted until the proportion of taxa in each AMBI EG corresponded to the expanded normative definitions for each ecological status (Figure 5.4). Each step is detailed below.

Identify pressure gradient data for EUNIS A5.3 habitat (Garroch Head sewage sludge disposal ground).

Calculate $IQI_{v.II}$ for macrobenthic samples over the defined pressure gradient.

Set equidistant boundaries within the $IQI_{v.II}$ range for each status class.

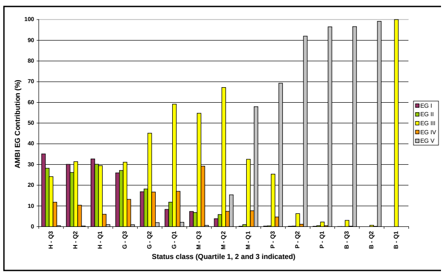
Boundary:	IQI value*
High/Good	0.8
Good/Moderate	0.6
Moderate/Poor	0.4
Poor/Bad	0.2

* IQI values equal to status boundaries will result in the upper status being assigned.

Assess AMBI ecological group proportions for each status class.

Adjust boundaries in the direction of the upper/lower quartile (that is, that reflecting closer resemblance to status described by the expanded normative definitions).

Calculate AMBI EG proportions for lower (1st) and upper (3rd) quartile of draft status classes.



Does the taxa composition in each status class comply with the expanded normative definitions?

No

Yes

Continue to test status class ranges on further datasets of the same habitat type.

Figure 5.4 Summary of the stepwise process used to establish boundaries for each status class to comply with the expanded normative definitions

The basis of the boundary setting approach was to adapt the boundaries so that each status achieved compliance with the expected abundance composition of AMBI ecological groups within the expanded normative definitions (Table 5.2).

The Garroch Head sewage sludge disposal ground dataset was used for the initial national boundary setting process, providing a broad pressure gradient over which samples from all status classes were represented. The EQR (using $IQI_{v.II}$) was calculated for each sample along with the percentage contribution of each AMBI ecological group.

Initially, the AMBI ecological group proportions were calculated for each status class for equidistantly set EQR boundaries (that is, 0.2, 0.4, 0.6, 0.8) to compare with the expanded normative definitions as a starting point for the iterative process (Figure 5.5).

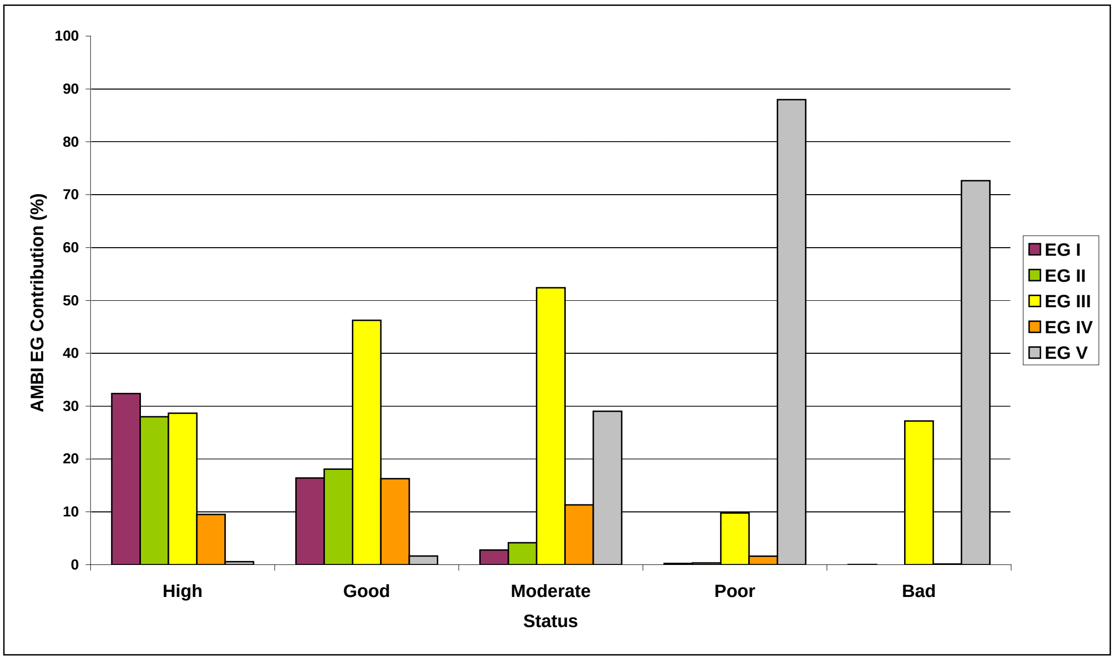


Figure 5.5 Contribution of AMBI ecological groups within each status class based on equidistant boundaries using the Garroch Head pressure gradient dataset (EUNIS A5.3)

However, based on equidistant class boundaries, the proportions of the five AMBI ecological groups for the different status classes were not consistent with the expanded normative definitions. For example, at moderate status, EGIII–EGV taxa are not co-dominant, high status is not dominated by EGI taxa, and so on. By analysing the AMBI ecological group proportions in the corresponding upper and lower quartiles of each class status, it was possible to identify whether a boundary increase/decrease would result in the AMBI ecological group proportions within each status better representing the expanded normative definitions. The AMBI ecological group proportions within the upper and lower quartiles from the equidistant class boundaries are presented in Figure 5.6.

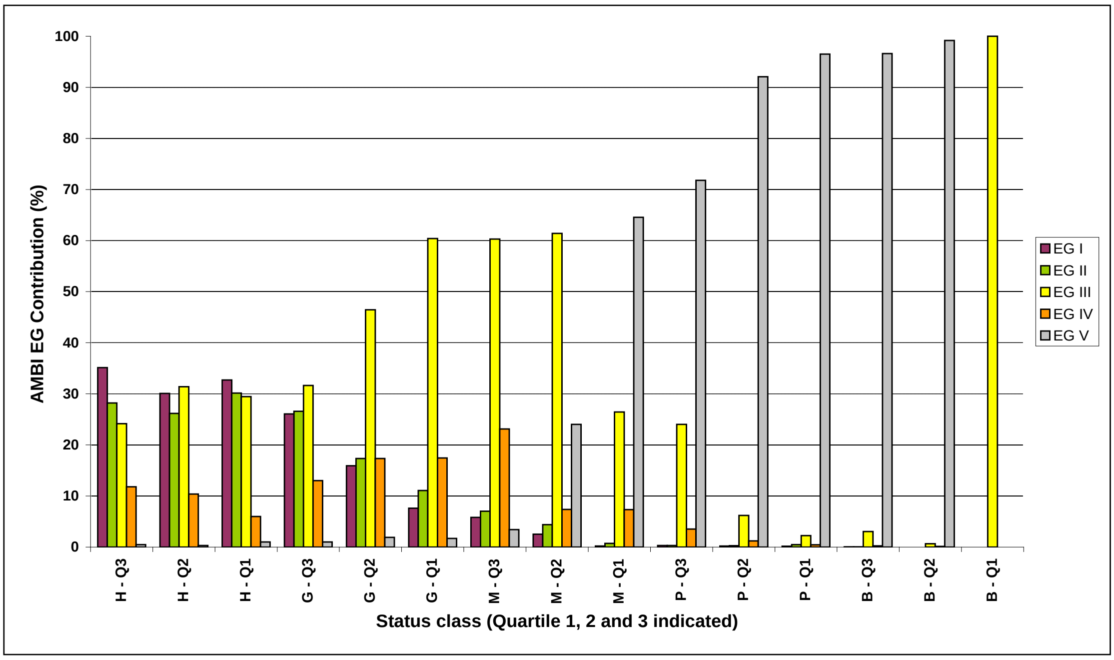


Figure 5.6 Contribution of abundance of different AMBI ecological groups within the lower quartile (Q1), median (Q2) and upper quartiles (Q3) of each ecological status class with class boundaries divided equidistantly

Where the AMBI ecological group proportions of the upper/lower quartile better represented the expanded normative definitions than the median, the upper/lower boundary was increased/decreased by an increment of 0.01 units on the EQR scale. This was repeated until the AMBI ecological group proportions within the median provided a greater alignment with the expanded normative definitions than those within the upper/lower quartiles (Figure 5.7).

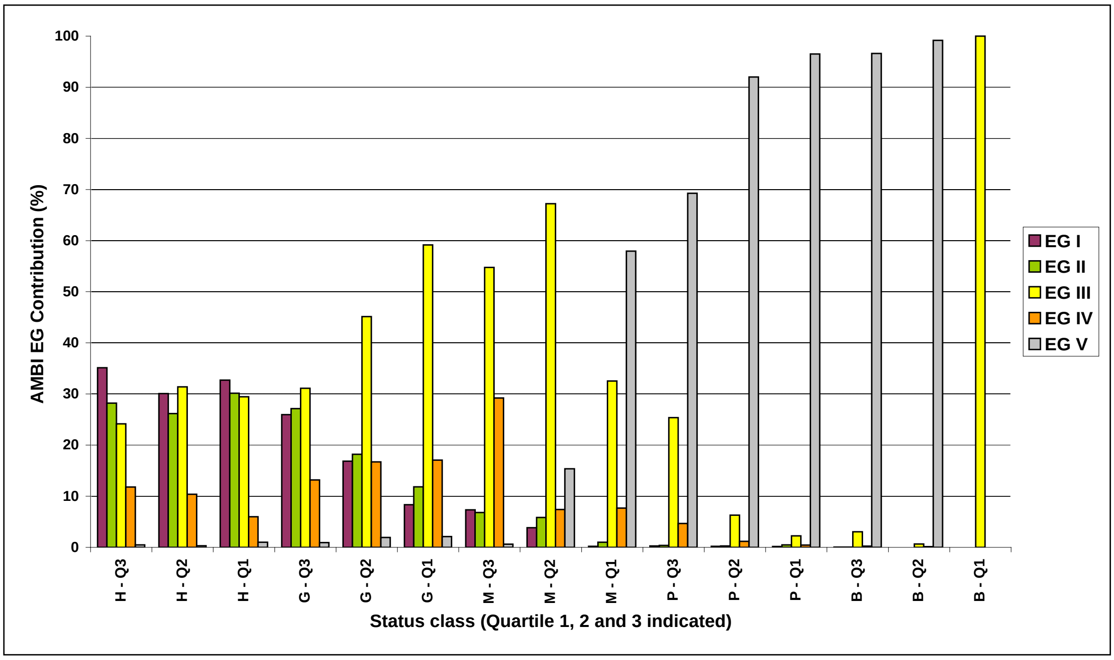


Figure 5.7 Contribution of abundance of different AMBI ecological groups within the lower quartile (Q1), median (Q2) and upper quartiles (Q3) of each ecological status class with class boundaries adapted to comply with the expanded normative definitions

Comparison between AMBI ecological group proportions from the equidistant boundaries and those adapted to fit the expanded normative definitions indicate only minor changes, indicating that equidistant boundaries provided a close agreement to the expanded normative definitions.

The class status EQR boundaries that provided the highest agreement between the median AMBI ecological group proportions and the expanded normative definitions are shown in Table 5.3. Note: the ecological class is assigned on values equal to or greater than the corresponding status boundary. The AMBI ecological group proportions for each class status based on the optimised boundaries are illustrated in Figure 5.8.

Table 5.3 Ecological status boundaries derived to provide the closest agreement between the median ecological sensitivity proportions and the expanded normative definitions

Status boundary	EQR
High–Good	0.80
Good–Moderate	0.65
Moderate–Poor	0.43
Poor–Bad	0.20

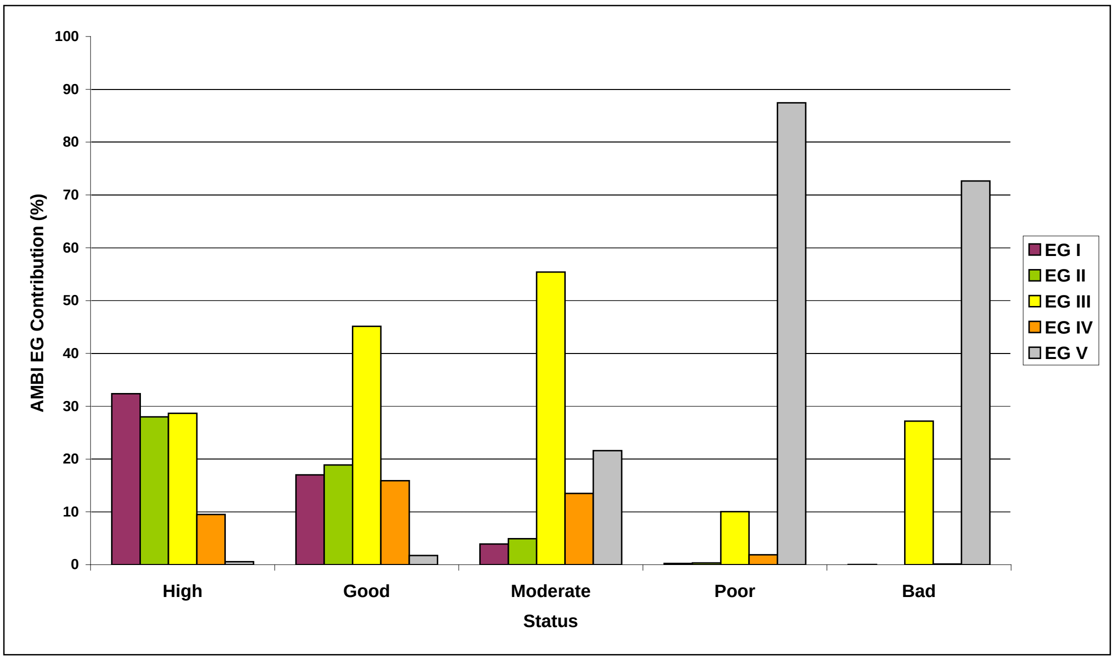


Figure 5.8 Contribution of AMBI ecological groups within each status class using derived boundaries shown in Table 5.3

These national boundaries were then taken forward to be adapted and finalised for coastal water types through Phase I of the intercalibration process (Section 5.3).

5.3 The intercalibration process

Section 1.4.1 of Annex V of the WFD on the comparability of biological monitoring methods sets a requirement that any proposed ecological assessment methodology must be intercalibrated between the Member States of each of the European maritime ecoregions: Baltic, North East Atlantic and Mediterranean. This process requires that the moderate/good and good/high boundaries are consistent with the normative definitions and comparable between Member States. Phase I Intercalibration (2004-2007) involved the comparison of $IQI_{v.II}$ and $IQI_{v.IV}$, and nationally proposed status boundaries for coastal waters, to the boundaries set by other members of the NEAGIG. The boundary optimisation process for the IQI involved only those assessment methods for a restricted section of North East Atlantic (NEA) ecoregion coastal water types: NEA1/26 and NEA7 (Table 5.4).

Table 5.4 North East Atlantic ecoregion water body types considered in Phase I intercalibration for setting Member State class status boundaries

Type	Salinity	Tidal range	Depth	Current velocity	Exposure	Mixing
CW-NEA1/26	Fully saline	Mesotidal (1–5 m)	Shallow (<30 m)	Low–medium (1–3 knots)	Exposed/ sheltered	Fully mixed
CW-NEA7 [1]	Fully saline	Mesotidal (1–5 m)	Deep (>30 m)	Low (<1 knot)	Sheltered	Fully mixed

Note: [1] NEA7 is applicable to Norway and the UK only.

The approach to the Phase I intercalibration of benthic ecological status assessment methods for coastal waters is described in Borja et al. (2007) and can be summarised in the steps shown in Figure 5.9.

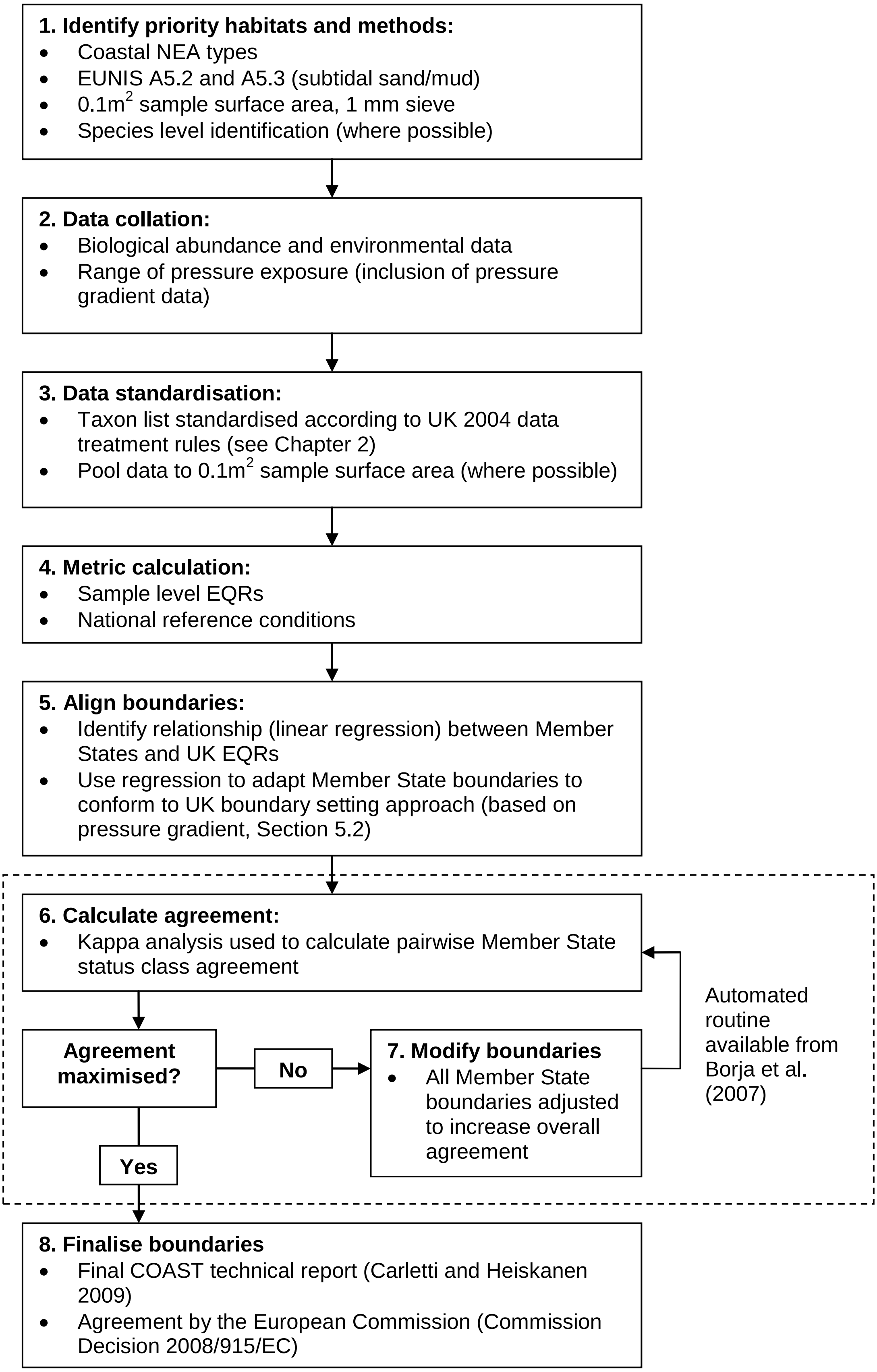


Figure 5.9 Overview of the Phase I intercalibration process for the setting of class status boundaries for coastal water types

5.3.1 Identification of common methodologies and priority habitats

To ensure that all Member States were able to use the data for method comparison, common sample collection and processing methods were identified. To minimise the variability which may result from different natural physicochemical conditions, Member States agreed that the boundaries should be intercalibrated from a limited range of habitats. EUNIS type A5.2 and A5.3 habitats (marine sublittoral sand and mud respectively) were selected on the basis that these habitats were considered to:

- be relatively stable with low degrees of natural pressure and variability (and consequently a higher likelihood of detecting a change in the benthic indicators as a response to an anthropogenic pressure signal)
- have relatively large amounts of available data due to historical contaminant monitoring focusing on these habitats

5.3.2 Data collation

Data (benthic infaunal abundance and supporting physicochemical parameters) were collated from all Member States (Table 5.5) to form a central NEAGIG intercalibration coastal water dataset consisting of 589 individual samples. To effectively intercalibrate all class status boundaries, it was necessary to include data exposed to varying degrees of anthropogenic pressure (for example, pressure gradient data).

Table 5.5 Summary of data provided by NEAGIG Member States for the Phase I intercalibration of coastal waters

Member State	Number of samples	Data features
Belgium	132	
Denmark	72	Includes eutrophication and hypoxia pressure data
Germany	64	
Ireland	14	
Norway	12	Includes hypoxia pressure and undisturbed data
Spain	45	Includes urban and industrial discharge pressure data
UK	250	Includes organic discharge pressure data

5.3.3 Data standardisation

The treatment of taxa at the laboratory analysis and data recording/storage stage varied between Member States. Variations included the taxonomic level to which certain taxa are identified (for example, class Oligochaeta), the inclusion/exclusion of non-benthic invertebrate taxa (for example, fish, algae, epifauna, pelagic taxa) and the enumeration of colonial taxa. Before use, the UK 2004 data treatment rules (see Chapter 2) were applied to the intercalibration dataset. Where data were available from methods of <0.1 m^2, replicates were pooled to achieve combined sample areas of ~0.1 m^2.

5.3.4 Calculation of metrics

The metrics for assessing coastal waters with corresponding reference conditions proposed by each Member State (Table 5.6) were applied to the standardised benthic dataset.[10]

Table 5.6 Summary of coastal water classification methods used by NEAGIG Member States during Phase I intercalibration

<table>
<tr><th>Member State</th><th>Method</th><th>Component metrics</th><th>Reference condition values</th></tr>
<tr><td>Denmark</td><td>Danish Quality Index (DKI) (see Borja et al. 2007)</td><td>AMBI (pollution sensitivity)
Shannon–Wiener (diversity)
Total abundance
Taxon number</td><td>AMBI = 0
Shannon-Wiener ($\log_2$) = 5
(Borja et al. 2007)</td></tr>
<tr><td>France</td><td>Multivariate factorial analysis (M-AMBI) (after Muxika et al. 2005)</td><td>AMBI (pollution sensitivity)
Shannon-Wiener (diversity)
Taxon number</td><td>AMBI = 1
Shannon-Wiener ($\log_2$) = 5
Taxon number = 68</td></tr>
<tr><td>Germany</td><td>Multivariate factorial analysis (M-AMBI) (after Muxika et al. 2005)</td><td>AMBI (pollution sensitivity)
Shannon-Wiener (diversity)
Taxon number</td><td><table><tr><td></td><td>NEA1</td><td>NEA26</td></tr><tr><td>AMBI</td><td>0.107</td><td>0.393</td></tr><tr><td>Shannon–Wiener ($\log_2$)</td><td>2.66
31</td><td>2.22</td></tr><tr><td>Taxon number</td><td></td><td>17</td></tr></table></td></tr>
<tr><td>Ireland</td><td>Infaunal Quality Index ($IQI_{v.II}$)</td><td>AMBI (pollution sensitivity)
Simpson's (evenness)
Taxon number</td><td>1-(AMBI/7) = 0.96
1-λ′ = 0.97
Taxon number = 68
(Section 4.4)</td></tr>
<tr><td>Norway</td><td>Norwegian Quality Index (NKI)</td><td>ISI (pollution sensitivity)
ES100 (diversity)
Taxon number</td><td>ISI = 9.9
ES100 = 26
Taxon number = 30
(Solheim et al. 2004)</td></tr>
<tr><td>Portugal</td><td>Portuguese Benthic</td><td>AMBI (pollution sensitivity)</td><td>AMBI = 0</td></tr>
</table>

[10] $IQI_{v.II}$ was used for the majority of the Phase I intercalibration process. However, the boundaries as included within Commission Decision 2008/915/EC were based on adaptations of the final boundaries to ensure maximum agreement with $IQI_{v.IV}$ following changes to the IQI during the Phase I process.

Member State	Method	Component metrics	Reference condition values
	Assessment Tool (P-BAT) [1]	Shannon–Wiener (diversity) Margalef (diversity)	Shannon–Wiener (log_2) = 4.1 Margalef = 5
Spain	Multivariate factorial analysis (M-AMBI) (after Muxika et al. 2005)	AMBI (pollution sensitivity) Shannon–Wiener (diversity) Taxon number	AMBI = 1 Shannon–Wiener (log_2) = 4 Taxon number = 42
UK	Infaunal Quality Index ($IQI_{v.II}$).	AMBI (pollution sensitivity) Simpson's (evenness) Taxon number	1-(AMBI/7) = 0.96 1-λ′ = 0.97 Taxon number = 68 (Section 4.4)

Note: [1] The Portuguese Benthic Assessment Tool (P-BAT) was developed during the course of the Phase I intercalibration and as such was not included in the early data analyses during the process.

Only a limited range of reference conditions was determined during the Phase I intercalibration. For the M-AMBI (after Muxika et al. 2005), as used by Spain, reference conditions were as described by Bald et al. (2005). Reference conditions had been partially established for the Norwegian Quality Index (NKI) (see Solheim et al. 2004). Approximate reference conditions were set for Shannon–Wiener for the Danish Quality Index (DKI). $IQI_{v.II}$ used the maximum metric values from UK pressure gradient data (Table 4.4).

Note: the approach to setting reference conditions has been developed further for each Member State since Phase I intercalibration.

5.3.5 Alignment of boundaries

The correlation between classification methods was calculated to identify whether the methods were compatible for Intercalibration. Pair-wise comparisons were made between all methods, with the degree of linear correlation being expressed by the pair-wise Pearson R^2 values (Table 5.7).

Table 5.7 Correlation coefficients (Pearson R^2) from linear regression between each pair of Member State assessment methods

Method	NKI	M-AMBI	$IQI_{v.II}$
DKI	0.69	0.84	0.83
$IQI_{v.II}$	0.67	0.68	–
M-AMBI	0.63	–	–

The boundaries originally proposed by each Member State (Table 5.8) had been developed independently and, as a result, differences existed in the approaches taken. Denmark separated the DKI scale into equidistantly spaced status classes. The approach adopted by Spain (Basque Region) for the M-AMBI (Borja et al. 2004) was based on recommendations from the WFD CIS Reference Conditions Working Group, REFCOND (2003). The approach set the good/high boundary according to the upper 10th percentile value within a dataset (set at 0.83), with the remaining scale being split equidistantly to derive the remaining status boundaries. The approach used by Norway in setting NKI boundary values is described by Solheim et al. (2004). The UK approach involved adapting class status boundaries until the AMBI ecological groupings within each status provided the closest agreements with the expanded normative definitions (Section 5.2).

Table 5.8 Original Member State ecological status boundaries proposed for each classification method

Method	Bad/poor boundary	Poor/moderate boundary	Moderate/good boundary	Good/high boundary
DKI	0.20	0.40	0.60	0.80
$IQI_{v.II}$	0.20	0.43	0.65	0.80
M-AMBI	0.20	0.41	0.62	0.83
NKI	0.47	0.60	0.72	0.83

As the UK approach determined boundaries according to an anthropogenic pressure gradient as required by the WFD, all Member State boundaries were adapted to align with these to form the basis of the Intercalibration boundary setting process. The regression formula from the linear relationship between the UK EQRs and those of other Member States was used to align the Member State boundaries with the UK boundaries accordingly (Table 5.9).

Table 5.9 Member State ecological status boundaries following alignment with UK boundaries

Method	Bad/poor boundary	Poor/moderate boundary	Moderate/good boundary	Good/high boundary
DKI	0.18	0.40	0.61	0.75
$IQI_{v.II}$	0.20	0.43	0.65	0.80
M-AMBI	0.16	0.42	0.67	0.84
NKI	0.25	0.46	0.66	0.80

5.3.6 Class status agreement:

The extent to which samples were in agreement in terms of status class was quantified using Kappa analysis (Cohen 1960, Landis and Koch 1977) with Fleiss–Cohen weighting (Fleiss and Cohen 1973) applied to acknowledge the increasing importance of disagreement by a greater number of status classes. The extent of agreement was also assigned qualitatively according to the methods of Monserud and Leemans (1992). The results of the Kappa analysis between Member State methods using the originally proposed national boundaries are presented in Table 5.10.

Table 5.10 Kappa agreement coefficients between Member State methods using status boundaries (aligned to UK boundaries) with corresponding Monserud and Leemans agreement and percentage of samples mismatched on either side of the moderate-good boundary

	NKI	M-AMBI	$IQI_{v.II}$
DKI	0.62 (good) 32.8%	0.84 (very good) 13.6%	0.81 (very good) 18.2%
$IQI_{v.II}$	0.62 (good) 26.7%	0.80 (very good) 21.2%	–
M-AMBI	0.61 (good) 31.7%	–	–

While the extent of agreement in assigned status class for the different methods ranged between 0.61 (good) and 0.84 (very good), optimisation of the boundaries was necessary in order to intercalibrate successfully by achieving maximum overall agreement.

5.3.7 Optimisation of boundaries

The boundaries established during alignment with those established by the UK (Table 5.9) were modified incrementally until maximum pair-wise correlations between all Member States were achieved. This was achieved using an automated Microsoft® Excel routine developed by Borja et al. (2007), which is available on request from the authors.

The approach to the intercalibration of ecological status boundaries described by Borja et al. (2007) related to an intermediate stage of Phase I intercalibration. During the intercalibration exercise further developments were made to the classification methods of several of the NEAGIG Member States. The IQI was revised from $IQI_{v.II}$ to $IQI_{v.IV}$ (see Chapter 3) and the reference conditions for EUNIS habitats A5.2 and A5.3 (0.1 m^2 grab processed to 1,000 µm) were updated (see Chapter 4). These developments had implications for the intercalibrated boundaries. IQI restructuring changed the class status assigned to the corresponding EQR values. The decrease in reference condition values for all metrics following the 2006 revision (Table 4.4), resulting in a general increase in the EQR value for all samples.

Without the adaptation of the status boundaries to accommodate these revisions, the agreement between the assigned status classes of the IQI and other NEAGIG methods decreased. In order to maintain the same level of agreement, the IQI boundaries from the optimisation process (Table 5.8) were readjusted to ensure the number of data values in each status class remained the same compared with the optimised boundaries with the original reference conditions.

In addition to the revisions to the methods initially proposed by the NEAGIG Member States, the Portuguese assessment method P-BAT was introduced. The addition of an extra method to the boundary optimisation process was likely to yield an overall sub-optimal agreement between all methods. As a result, the optimisation process was repeated. The process resulted in the finalisation of all status class boundaries for $IQI_{v.IV}$ for coastal waters (Table 5.11).

Table 5.11 Ecological status boundaries proposed for North East Atlantic coastal types NEA1/26 and NEA7 following the optimisation process

Method	Moderate/good boundary	Good/high boundary
DKI	0.53	0.67
$IQI_{v.IV}$	0.64	0.75
M-AMBI	0.53	0.77
NKI	0.81	0.92
P-BAT	0.58	0.79

Improvement in agreement between class statuses was observed in all pair-wise comparisons following the updates to the optimised boundaries, with increases in correlation coefficients ranging between 0.01 (DKI and M-AMBI) and 0.21 (M-AMBI and NKI) (Table 5.12).

Table 5.12 Kappa agreement coefficients between Member State methods using optimised status boundaries with corresponding Monserud and Leemans agreement and percentage of samples mismatched

	P-BAT	NKI	M–AMBI	$IQI_{v.IV}$
DKI	0.89 (almost perfect) 10.59%	0.91 (almost perfect) 7.91%	0.93 (almost perfect) 5.51%	0.91 (almost perfect) 8.47%
$IQI_{v.IV}$	0.85 (almost perfect) 13.8%	0.96 (almost perfect) 3.11%	0.86 (almost perfect) 12.53%	–
M-AMBI	0.92 (almost perfect) 6.62%	0.87 (almost perfect) 11.72%	–	–
NKI	0.87 (almost perfect) 12.29%	–	–	–

The implications of the revised boundaries in terms of the relevance to the AMBI ecological groups over the EQR scale can be observed in Figure 5.10.

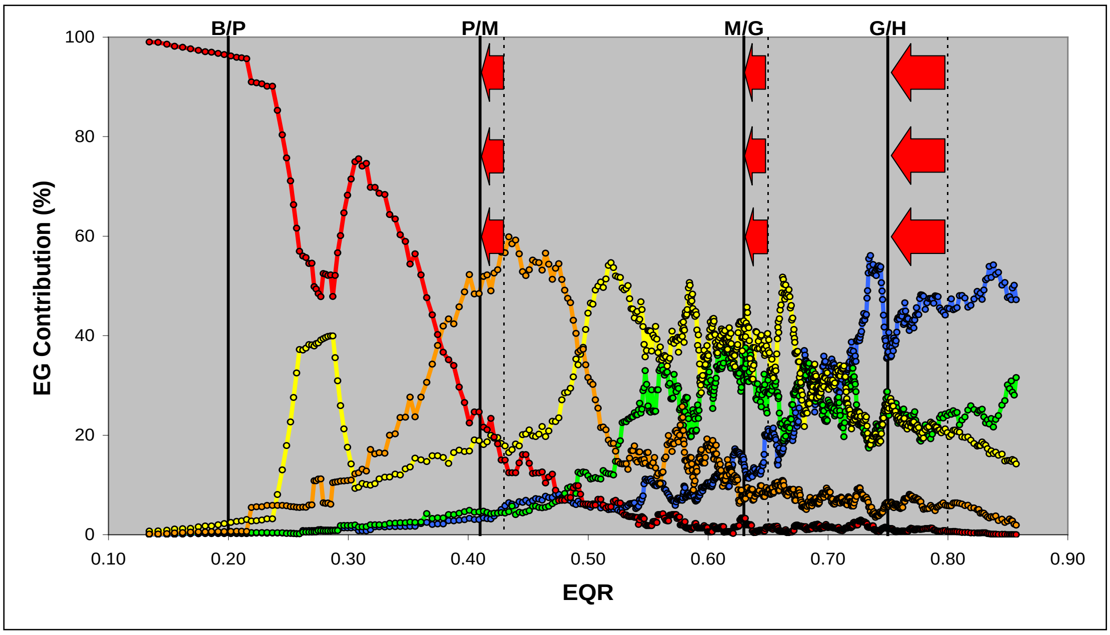


Figure 5.10 Proportion of AMBI ecological groups over the EQR scale within the North East Atlantic Phase I intercalibration data indicating positions of nationally derived status boundaries (dashed lines) and status boundaries following optimisation in intercalibration Phase I (solid lines)

Notes: B/P = bad/poor boundary

P/M = poor/moderate boundary

M/G = moderate/good boundary

G/H = good/high boundary

Individual points represent average EG proportions from multiple samples to illustrate the underlying pattern.

The final class status boundaries determined during Phase I Intercalibration were presented in the Intercalibration COAST technical report and accepted by the European Commission in 2008 (European Commission 2008) (Table 5.13).
The boundaries for the M-AMBI for Germany were revised from those recommended through the intercalibration process on the basis of expert judgment by the Member State.

Table 5.13 Finalised class status boundaries as established during Phase I intercalibration as accepted by the European Commission

Type and country	National classification system	Ecological Quality Ratios	
		High-Good boundary	Good-Moderate boundary
Types NEA1/26, NEA 3/4 and NEA7 (Indices responsive primarily to organic enrichment and toxic pollution pressures in soft sediment habitats)			
Denmark	DKI	0,67	0,53
France	M-AMBI	0,77	0,53
Germany	M-AMBI	0,85	0,70
Ireland	IQI	0,75	0,64
Norway	NQI	0,92	0,81
Portugal	P-BAT	0,79	0,58
Spain	M-AMBI	0,77	0,53
United Kingdom	IQI	0,75	0,64
Types NEA1/26 and NEA3/4 (Index responsive to multiple pressures in multiple habitats)			
Belgium	BEQI	0,80	0,60
Netherlands	BEQI	0,80	0,60
Types NEA8/9/10			
Denmark	DKI	0,82	0,63
Norway	NQI	0,92	0,81
Sweden	BQI	0,89	0,68

Source: Commission Decision 2008/915/EC (European Commission 2008)

5.4 Transitional water class boundaries

The inclusion of reference conditions within $IQI_{v.IV}$ has enabled the coastal water boundaries to be applied. Formal setting of transitional water status class boundaries through intercalibration (type NEA11) is the focus of Phase II.

6 Variability

WFD assessments are estimates of the status of the ecological quality elements within an entire water body. It is necessary to define the precision of the estimates to establish the statistical confidence of a water body assessment (confidence of class) and to establish sample effort for monitoring programmes (power analysis). This chapter describes how assessments are affected by different aspects of variability and the approach adopted in calculating variability for WFD water body assessments.

6.1 Introduction

WFD Annex V, section 1.3 on the monitoring of ecological and chemical status for surface waters states that:

> 'Member States shall monitor parameters which are indicative of the status of each relevant quality element. In selecting parameters for biological quality elements Member States shall identify the appropriate taxonomic level required to achieve adequate confidence and precision in the classification of the quality elements. Estimates of the level of confidence and precision of the results provided by the monitoring programmes shall be given in the plan'.

Annex V, section 1.3.4 on the frequency of monitoring states that:

> 'Frequencies shall be chosen so as to achieve an acceptable level of confidence and precision. Estimates of the confidence and precision attained by the monitoring system used shall be stated in the river basin management plan'.

Macroinvertebrate water body assessments are based on quantitative sample data, which provide only a small representation of the overall benthic infaunal population. For soft sediment infauna, it is not possible to obtain information on the full invertebrate population of a water body due to the need to process sediments in order to obtain information on the macrobenthic infaunal community. (For some quality elements, remote imagery such as aerial photography may provide information about the full extent or population for the water body.)

Sampling exposes assessments to sampling error and, by basing an assessment on sampling, estimates of ecological quality are likely to be an approximation of the true value in the underlying population (Ellis and Adriaenssens 2006). Sampling error may result in a water body being assigned a different status class to its true class. This is termed the 'risk of misclassification' (RoM). RoM is used to fulfil the WFD requirement to report the precision in the classification of quality elements.

Monitoring programmes require that sample data to contain sufficient information for the classification to provide an acceptable level of precision or RoM. Variability information can be used to provide an estimation of the sampling effort (the spatial and temporal frequency of samples) required to attain the prescribed levels of precision. The probability of detecting differences in a hypothesis test can be estimated using power analysis. In a WFD context, this is the probability of detecting whether the EQR of a water body is different from the moderate/good boundary, or has achieved GES. Sampling effort is intrinsic to power analysis, and by investigating the implications of sampling effort through power analysis, the number of samples that are required to detect whether a water body mean EQR differs from the moderate/good boundary with a prescribed level of statistical certainty can be estimated.

Quantitative estimates of IQI sampling error or variability are key elements in the calculation of the RoM and for defining the power of a monitoring programme. Details of the approach to the RoM and power analysis methods applied to the IQI are given in Chapters 7 and 8 respectively. The process for deriving EQR variability estimates, along with the associations to estimating sampling effort through power analysis and statistical certainty of assessments through confidence of class (CofC) and risk of misclassification (RoM), is summarised in Figure 6.1.

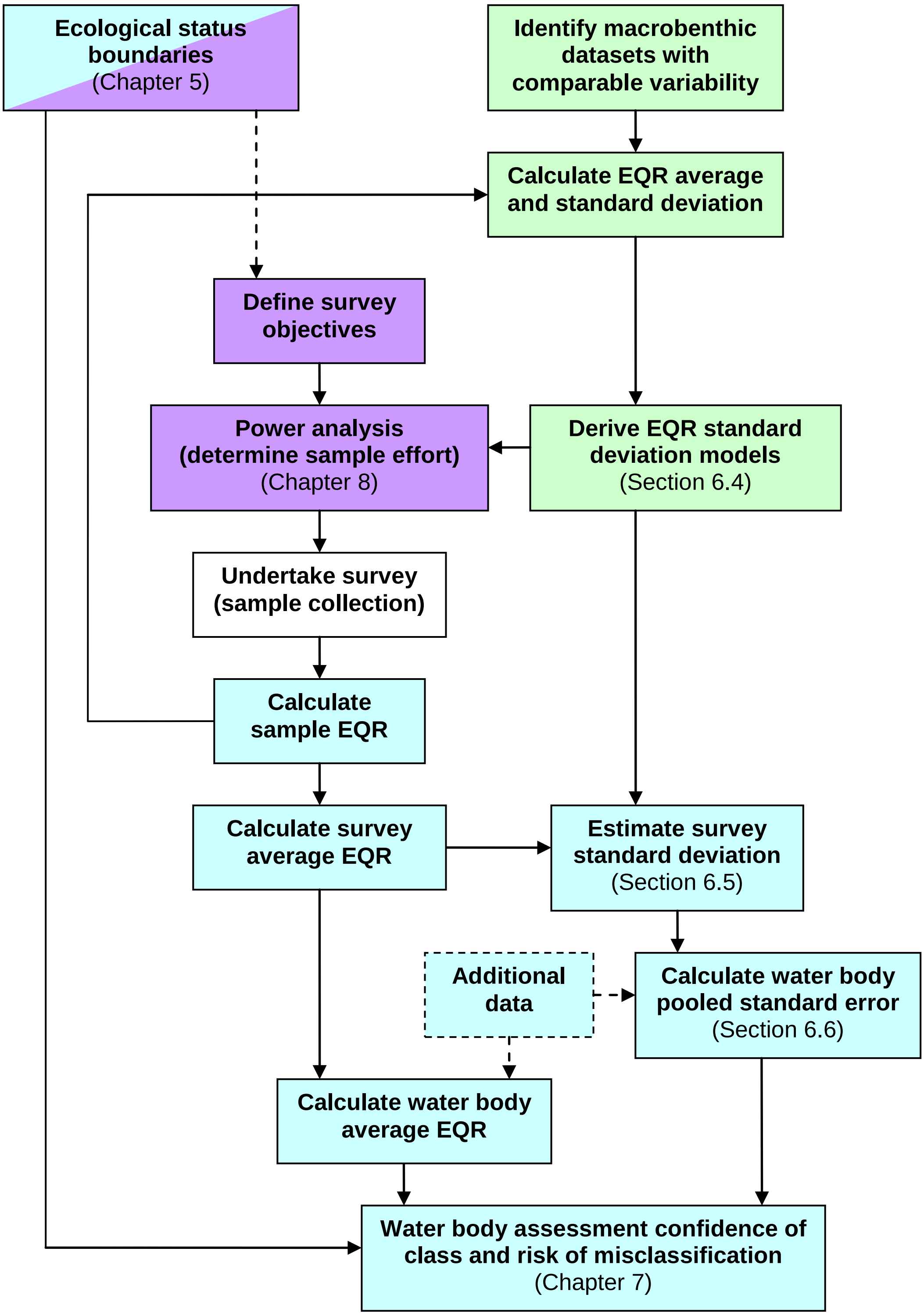


Figure 6.1 Main components in deriving statistical certainty of assessments

Notes: Green: process interactions between the components involved in deriving EQR variability estimates
Purple: estimating sampling effort through power analysis
Blue: confidence of class and risk of misclassification

6.2 Sources of EQR variability

Repeat measurements of the same parameter often provide differences in the observed results due to variability. In ecological systems these sources can be attributed to being natural or anthropogenic in origin. The variability in an observed value as an additive function of changes in the true value of the parameter being observed and error is described by classical test theory. This is summarised as:

$$X = T + e_s + e_r \qquad \text{Equation 6.1}$$

where:

X = observed value

T = true value

e_s = systematic error (bias)

e_r = random error (sample and measurement)

The principle assumes that a true value (T) exists (a true response to anthropogenic pressure in the current context) that would be evident if no error existed. Understanding systematic error (e_s) and reducing random error (e_r) allows T to be estimated with a greater degree of accuracy and precision respectively. The approach to estimating systematic error through the development of habitat specific reference conditions is described in Chapter 4.

Random error is used to describe the variability resulting from sampling and measurement error and ecological interactions (see below). It should not affect the estimated average EQR, only the departure of the individual data around the average. Systematic error (or bias) refers to the tendency to consistently under or overestimate a true value. It therefore has the potential to affect the average EQR (and corresponding status class) for a water body assessment.

The extent to which the true value of a measured parameter is influenced by random and systematic error is also referred to as the signal:noise ratio. Chapter 4 describes how the IQI attempts to increase the signal:noise ratio by minimising the systematic error (bias) in EQRs that result from changes in habitat by factoring physicochemical data into the IQI calculation. As a consequence, the IQI is susceptible to variability and bias in both the biological and the physicochemical data. Random error can be reduced by applying standardised field protocols and analytical quality control schemes to sample collection and analysis.

Random error and bias in the sampling of benthic invertebrate assemblages may result from one or more of the following sources of error.

6.2.1 Sampling and measurement error

Various factors interfere with the ability to measure the true value of a biological or physicochemical parameter, causing variability in the observed value.

For benthic infauna data, sampling error may arise from, for example, chance retention of fauna smaller than the sieve mesh or loss of fauna during processing. Potential sources of measurement error may occur from incorrect identification or enumeration, and use of approximations (for example, subsampling).

Examples of sources of sampling error for the supporting physicochemical data include differences in PSA values resulting from small-scale sediment heterogeneity and differences in salinity from the profile of an unmixed transitional water. Measurement error may result from factors such as poor instrumentation calibration.

By applying standardised field protocols and analytical quality control schemes, monitoring programmes aim to reduce error at the stages of sample collection and analysis.

In addition to error arising during the collection and measurement of the biota, variability also results from ecological processes such as predator/prey interactions, competition and alteration by natural disturbance events.

6.2.2 Temporal variability

Temporal variability may be apparent in:

- predictable patterns (for example, within-year changes in temperature)
- a combination of less predictable changes such as variations in certain taxa due to variation in predator populations or recruitment success
- possible long-term fluctuations

The effect of temporal variability may be reduced by defining a limited sampling window, for example, CSEMP restricts benthic infaunal sampling to between February and May inclusive.

6.2.3 Spatial variability

Differences between EQR values may occur across a water body due to variations in localised pressures or physicochemical conditions (for example, salinity changes across a transitional water body). As with temporal variability, these sources of variability may systematically affect EQR values (for example, changes over a water body due to a salinity or anthropogenic pressure gradient) or bring about apparently random changes.

6.2.4 Interaction variability

Temporal and spatial variability may not necessarily be independent, for example, aspects contributing to spatial variability may be linked to temporal factors. Such interactions include the effect of seasonal changes in freshwater flow into a transitional water on the spatial variability in salinity, or the frequency and timing of activities such as dredging in specific locations.

In cases where assessments are exposed to both temporal and spatial variability, this temporal–spatial interaction variability is ideally taken into account. However, interaction variability is often difficult to isolate and is therefore incorporated into the residual sample and measurement error.

Understanding how spatial and temporal variability affects the EQRs enables monitoring programmes to be developed to minimise their impact and to increase confidence in the corresponding assessment results.

6.3 Quantifying EQR variability

The approach to estimating EQR variability for the purpose of RoM was developed in collaboration with the Water Research Centre (WRc) and is described in detail in terms of its general application to WFD classifications by Ellis and Adriaenssens (2006).

EQR variability needs to be quantified for incorporation into the analysis of the risk of misclassification and the power analysis of monitoring programmes. Two options were considered to determine EQR variability:

- calculate variability directly for a given dataset
- estimate variability from existing data with comparable variability

In the case of the first option, the standard deviation calculated for a single dataset is considered an estimate of the true standard deviation and also has a degree of uncertainty. To minimise the effect of this inherent variability for standard deviation values from a single dataset, the second option was therefore adopted.

Ideally, components of variability would be established through surveys specifically designed to estimate such values in accordance with sound statistical principles (Ellis and Adriaenssens 2006). However, the ability of the competent authorities to adopt this approach is often limited by practicality and cost.

Because EQR variability for the power analysis must be considered at the survey design stage, the option of establishing variability directly for a given dataset is not possible so it must be estimated a priori. The expected variability must be therefore estimated from existing data with comparable characteristics in terms of factors influencing variability.

In establishing the Environment Agency's WFD surveillance monitoring programme, power analysis of historical data was undertaken to derive initial approximate sampling effort to classify a water body, to be revised for subsequent monitoring cycles once the first round of surveillance monitoring data had been collected.

6.4 Modelling EQR variability

The basis of the models used to estimate EQR standard deviations lies in the relationship between the mean and standard deviation of the EQRs. This relationship can be illustrated using benthic invertebrate data from ecological monitoring between 1980 and 1997 in the lower Tees estuary programme (AstraZeneca unpublished data). These data were collected over a period of 17 years when localised changes in industrial activity, salinity and tidal regime changes from the 1995 Tees barrier development and possible broad-scale environmental changes (Warwick et al. 2002) mean that the data represent a range in ecological condition. Thus the dataset provides a range of average EQR values over which the average EQR versus the standard deviation (SD) can be illustrated. Each station within the dataset holds multiple replicates (3–5) for multiple years on which the within-station mean EQR and associated standard deviation values (for a given year) were calculated (Figure 6.2).

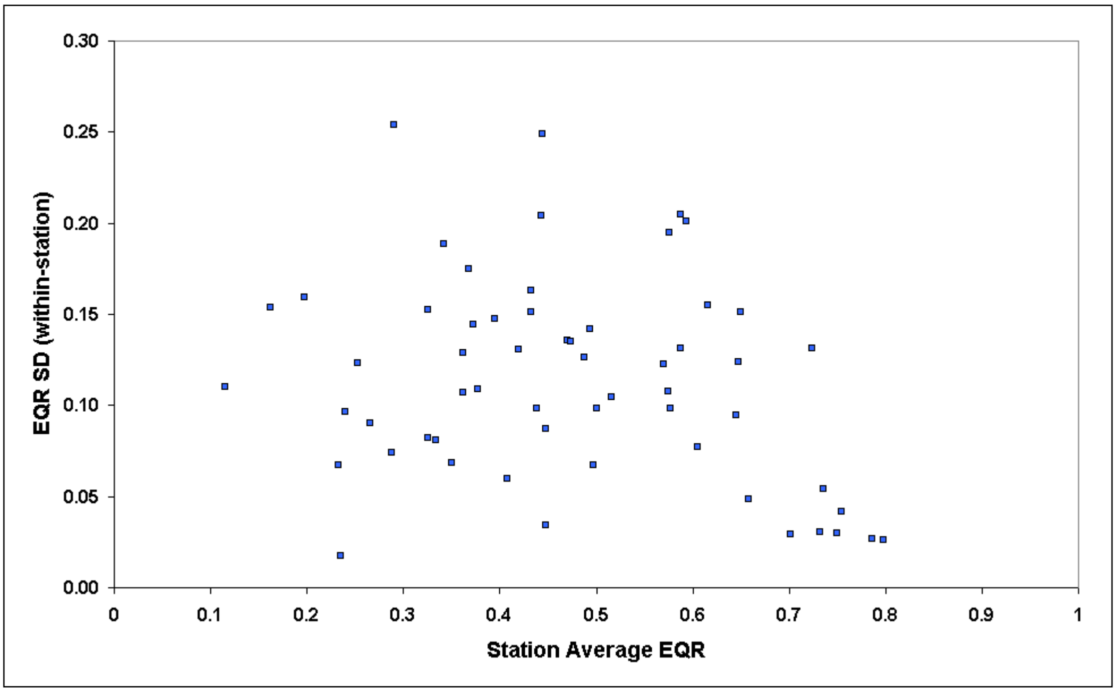


Figure 6.2 Station mean EQR versus within-station standard deviation EQR values for ecological monitoring in the lower Tees estuary data surveyed between 1980 and1997 (data provided by AstraZeneca)

In developing a suitable model to predict the EQR standard deviation using the average EQR, the following assumptions were applied:

- As the EQR has a lower and upper limit of zero and one respectively, variability is zero where the true EQR is equal to one or zero. The relationship between EQR average and standard deviation is not linear.
- Sampling error is normally distributed or approximately normal following transformation.

On this basis, the beta-curve model was proposed as a means of estimating the EQR standard deviation, defined in the following expression:

$$y = A(x^{p}) \times (1-x)^{q}$$ Equation 6.2

where:

x = station mean EQR

y = within-station standard deviation

A/p/q = beta-curve parameters estimated from the data

The beta-curve is characterised by two positive parameters, p and q, and is defined for values of x (in this case, average EQR) running between 0 and 1. The distribution becomes a parabolic expression where p = q = 2. The tails of the curve approach the x-axis with the peak of the curve lying anywhere between 0 and 1 according to the values of the parameters p and q.

By log transforming the data, the relationship between EQR average and standard deviation can be expressed as a linear function (rather than beta-curve) as follows:

$$Log_{10}(y) = \log_{10}(A) + (p \times \log_{10}(x)) + (q \times \log_{10}(1-x))$$ Equation 6.3

The expression of the relationship as a linear function enables the parameters p and q to be estimated by regression analysis between the predictor variables, $\log_{10}(x)$ and $\log_{10}(1-x)$, and the response variable, $\log_{10}(y)$. In this case, regression analysis was used as a curve fitting tool and not for hypothesis testing, so the assumption of multiple regression that transformations of the x-variable should be independent of each other was not valid and it was appropriate to use two transformations of the EQR to carry out the regression.

The relationship between the predictor variables ($\log_{10}(x)$ and $\log_{10}(1-x)$) and the response ($\log_{10}(y)$), was then identified through regression analysis to calculate the coefficient values to be used in the models (Table 6.1). The coefficients of $\log_{10}(x)$ and $\log_{10}(1-x)$ are equal to the slopes of the two x-variables. The values used in the model formula were A, which is the anti-log of the intercept, and the values of p and q, which are the x-variable coefficients.

Table 6.1 Coefficients from regression analysis between log10(standard deviation) (response variable) and $\log_{10}$(average) and $\log_{10}$(1-average) (predictor variables) with corresponding beta-curve model values using lower Tees data (1980-1997) from AstraZeneca

Linear regression coefficients		Beta-curve model values	
Intercept	0.25	A	1.78
$\mathrm{Log}_{10}(x)$	1.63	p	1.63
$\mathrm{Log}_{10}(1-x)$	0.55	q	0.55

These values were applied to Equation 6.2 to produce the beta-curve for the lower Tees data to estimate the EQR standard deviation (y) from the EQR average (x) over the zero to one scale to illustrate the fit of the beta-curve to the data (Figure 6.3).

There is a considerable degree of scatter of the individual data points around the model. This may result from between-station and between-year variations in environmental conditions that are influential to EQR variability. While the model is sufficient in estimating the standard deviation at a given EQR in the context of the lower Tees estuary data in general, the separation of data according to factors that influence variability would improve the accuracy of the models in estimating standard deviation from the mean. The fit of the beta-curve through the data indicates approximate normality of the points about the model.

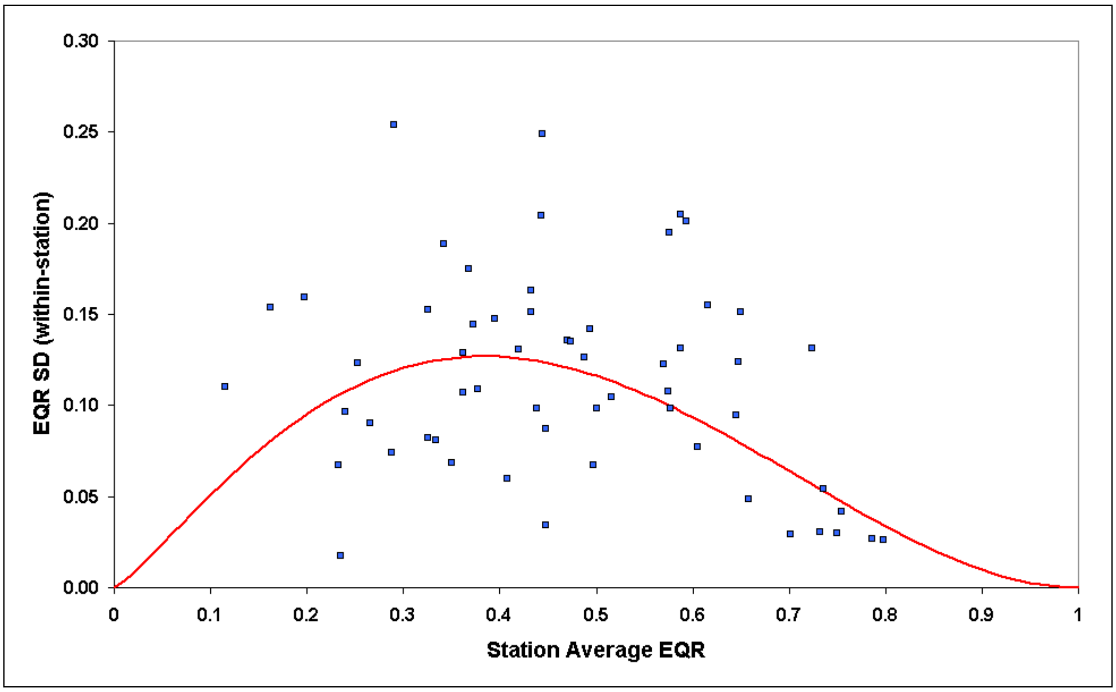


Figure 6.3 Station mean versus within-station standard deviation EQR values with fitted beta-curve model for ecological monitoring in the lower Tees estuary data surveyed between 1980 and 1997 (data provided by AstraZeneca)

The beta-curve models produced should be considered specific to the nature of the variability in the data on which they are based, that is, the expected components (Section 6.5) and the magnitude of the variability of the test data should be comparable with that of the data underlying the model. Inaccurate variability estimates are likely to result from estimating variability for a test dataset using a model based on:

- different components of variability (for example, where between-year variability is absent from the model data but present in the test data)
- data with different magnitudes of variability (for example, using between-station variability models from transitional water data to estimate variability for coastal water data)

The EQR values in the Tees example are an average of multiple samples taken on a single occasion at a single station and are therefore subject to both within-station variability and sampling and measurement variability.

6.5 Combining EQR variability estimates

The overall variability for a water body assessment depends on how the status of a water body is defined (Ellis and Adriaenssens 2006). For the purposes of defining the risk of misclassification, confidence of class and power analysis, it is necessary to calculate the overall variance of the assessment data. Where assessment data are subject to multiple components of variability, these must be combined to enable the overall variability or standard deviation to be calculated.

The overall standard deviation depends on the applicable temporal (for example, within/between year) and spatial (for example, within/between station) components of the variability of the sample data. These are estimated using the standard deviation beta-curve model approach as described in Section 6.4. The different components of variability can be used (along with the number of sites/replicates/sampling occasions) to calculate the overall variance as described by Ellis and Adriaenssens (2006):

$$\text{Variance} = \left(\frac{\text{SD E}^2 + \text{SD Int}^2 + \text{SD Local}^2}{\text{No. Occasions} \times \text{No. Sites} \times \text{No. Replicates}} \right) + \left(\frac{\text{SD Seas}^2 + \text{SD Temp}^2}{\text{No. Occasions}} \right) + \left(\frac{\text{SD Grad}^2 + \text{SD Spat}^2}{\text{No. Sites}} \right)$$

Equation 6.4

where:

SD E = standard deviation due to sampling error

SD Int = standard deviation due to spatial-temporal interaction

SD Local = standard deviation due to local spatial variability at a site

SD Seas = standard deviation due to seasonal cycles

SD Temp = standard deviation due to random temporal variation

SD Grad = standard deviation due to systematic spatial trend along a water body

SD Spat = standard deviation due to random spatial variability

The WRc developed a tool referred to as ‘CAVE’ (Combines Appropriate Variance Estimates) which calculates the overall variance of a dataset by combining the different components of variance in Equation 6.4. Further information on the CAVE calculation tool is available in Ellis and Adriaenssens (2006). Although the variance formula given above can be applied to a range of WFD quality element classification systems, Ellis and Adriaenssens (2006) emphasised that different classification tools should be accompanied by clear statements of how the formula should be applied for a given set of classification data.

6.6 Calculating standard error

The standard deviation values estimated as described in Sections 6.4 and 6.5 are used in the calculation of the standard error (SE) of the mean EQR. This forms the basis of the approaches to risk of misclassification, confidence of class (Chapter 7) and power analysis (Chapter 8).

WFD water body level assessments are based on the average EQR and may incorporate data from multiple sampling methods and habitats or contain data from multiple of sampling approaches (for example, single samples versus multiple replicates at each sampling location). With variability being partly dependent on such factors, to use sample data from a range of such conditions, the pooled standard error of the mean EQR is calculated as:

$$SE = \sqrt{\sum_{i=1}^{?} \left(\frac{SD_i^2}{n_i} \right)}$$

Equation 6.5

where:

SE = standard error

SD = standard deviation

n = number of samples

Reducing the overall standard error of the EQR data will increase the significance of an observed EQR being within a specified proximity to a specified threshold (therefore increasing CofC and reducing RoM). This can be achieved by a range of approaches:

- increasing sample numbers, that is, increasing values of n (Equation 6.5)
- excluding components of variability that do not assist in addressing the test hypothesis (for example, sampling a single location/occasion to avoid additional spatial/temporal variability)
- adopting sampling collection and processing methods that provide reduced measurement error (for example, larger sample surface areas, finer sieve mesh).
- reducing measurement error by applying appropriate sample collection, processing and analysis quality assurance procedures.

However, these actions must not be applied to the detriment of the ability of the data to address the test hypothesis. For example, the adaptation of a survey to exclude a temporal component of variability would be limited within a programme designed to detecting temporal trends. The cost–benefit implications of changes need to be considered at the survey design stage.

6.7 Variability and monitoring programmes

WFD monitoring requires that EU Member States establish monitoring programmes to classify water bodies into the five status classes. According to specific objectives (Box 6.1), monitoring programmes are to be undertaken for:

- surveillance
- operational
- investigative purposes

Box 6.1 Design of WFD monitoring for ecological status and chemical status for surface waters

The following objectives are set out in WFD Annex V.

Design of surveillance monitoring (section 1.3.1)
‘Member States shall establish surveillance monitoring programmes to provide information for:

- supplementing and validating the impact assessment procedure detailed in WFD Annex II,
- the efficient and effective design of future monitoring programmes,
- the assessment of long-term changes in natural conditions, and
- the assessment of long-term changes resulting from widespread anthropogenic activity.’

Design of operational monitoring (section 1.3.2)
‘Operational monitoring shall be undertaken in order to:

- establish the status of those bodies identified as being at risk of failing to meet their environmental objectives, and
- assess any changes in the status of such bodies resulting from the programmes of measures.

Investigative monitoring (section 1.3.3)
'Investigative monitoring shall be carried out:

– where the reason for any exceedence is unknown,
– where surveillance monitoring indicates that the objectives set out in Article 4 for a body of water are not likely to be achieved and operational monitoring has not already been established, in order to ascertain the causes of a water body or water bodies failing to achieve the environmental objectives, or
– to ascertain the magnitude and impacts of accidental pollution.

and shall inform the establishment of a programme of measures for the achievement of the environmental objectives and specific measures necessary to remedy the effects of accidental pollution.'

The implications of variability on the monitoring programmes are described in this report within the context of the Environment Agency surveillance monitoring programme.

Between 2007 and 2009, benthic invertebrate sampling for surveillance monitoring was undertaken in the coastal and transitional waters of England and Wales. Between 15 and 45 benthic invertebrate samples, with paired physicochemical data, were taken from 43 water bodies (17 coastal water bodies and 26 transitional water bodies). The approach to surveillance monitoring was developed prior to the development of habitat-specific IQI reference conditions.

Initial analysis of historic data from a variety of sources indicated increasing variability with water body size. Consequentially, sample effort was modified according to water body size with either 15, 30 or 45 samples being taken for small (<4km^2 for transitional waters, <40km^2 for coastal waters), medium (4–40km^2 for transitional waters, 40–400km^2 for coastal waters) and large (>40km^2 for transitional waters, >400km^2 for coastal waters) water bodies respectively. The increased IQI variability with water body area observed within the initial analysis may have been a consequence of using static IQI reference conditions (that is, changes in habitat were not accommodated) with variability being driven in part by increasing habitat heterogeneity with increasing water body area (of particular relevance to changing salinity in transitional waters).

The Environment Agency's WFD surveillance programme followed the approach of monitoring single samples, evenly dispersed (spatially) across suitable substrata within the water body, on a single sampling occasion. In general, as a result of the sampling methods currently used and the available data on which valid reference conditions were based, sampling was limited to a restricted range of sediment types. This approach attempted to minimise the overall variability of the WFD assessments by minimising the components of variance to:

- sampling and measurement variability
- within-station (water body) variability

Other components such as between-station and between-year do not therefore apply. This allowed the overall variability of a water body assessment to be expressed as the standard deviation of all samples about the mean.

Further details of the Environment Agency surveillance monitoring programme are documented in its Operational Instruction for WFD macrobenthic sampling in transitional and coastal waters (Environment Agency 2012).

Data from the first stage of the Environment Agency surveillance monitoring (2007-2009) were used to refine the surveillance programme for second stage of monitoring (2010-2012). The inclusion of habitat-specific reference conditions in the IQI took place after the design of the first stage of surveillance monitoring. As the inclusion of reference conditions attempts to reduce the bias in EQR values that results from changing habitats and considered to be associated with spatial changes within a water body (particularly over salinity gradients within transitional water bodies), this had implications on the approach to monitoring. This potentially allowed the sample effort of the monitoring programme to be revised to be dissociated from water body size. Instead, as such, the 2007-2009 WFD data were analysed to address the following hypotheses.

- There is no correlation between EQR standard deviation and water body area.
- There is no difference between EQR standard deviation for different water body categories (coastal waters and transitional waters) and tidal exposure (intertidal and subtidal).

Regression analysis was used to test for linear correlations between EQR standard deviation and water body area ($\log_n(m^2)$). As transitional water bodies generally have a smaller surface area than coastal water bodies (that is, salinity is indirectly linked to water body area), the water body categories were analysed separately to avoid any correlation between variability and salinity being interpreted as a correlation between water body area and variability. Likewise, intertidal and subtidal data were analysed separately on the basis that water body area and sampling method may not be independent as small water bodies are often sampled intertidally due to shallow water depth limiting survey vessel access.

In all combinations of water body category and sampling method, the correlation between water body size and EQR standard deviation as tested using regression analysis was not significant (p >0.05) (Figures 6.4 to 6.6, Table 6.2). This indicates that sampling effort for IQI assessments does not need adapting according to water body area. The correlation between water body area and EQR standard deviation was not tested for coastal water intertidal surveys due to insufficient data.

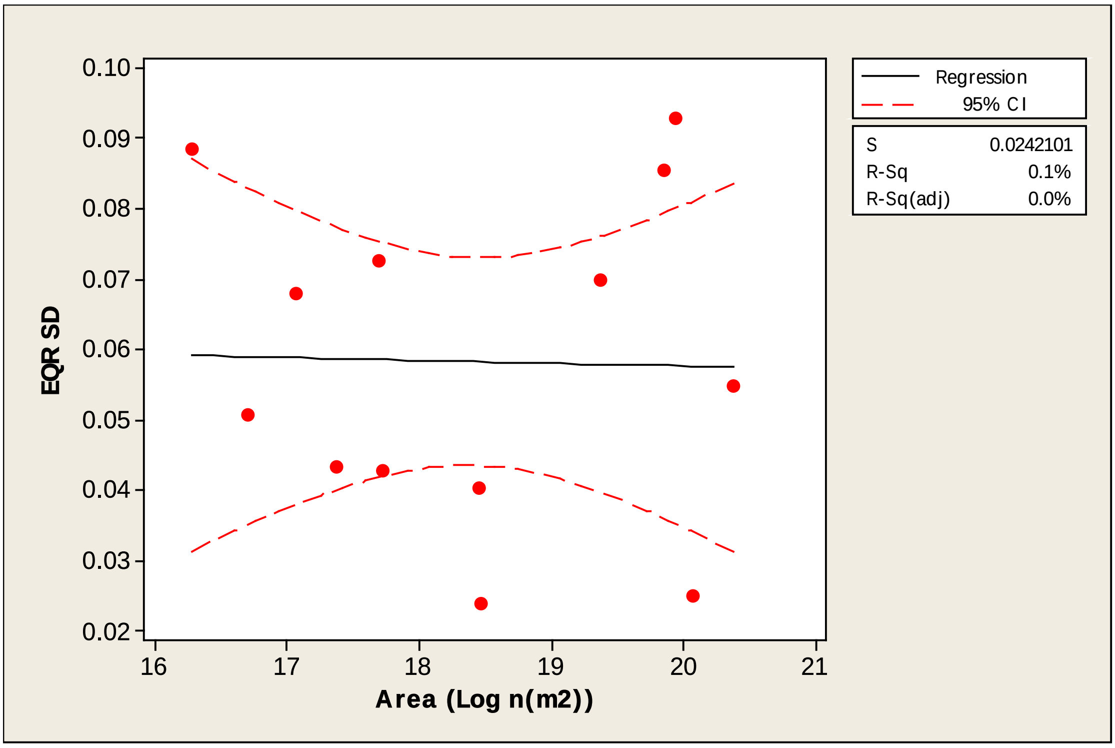


Figure 6.4 EQR standard deviation versus water body area ($\log_n(m^2)$) with upper and lower 95% confidence intervals (CIs) for coastal water subtidal WFD 2007-2009 surveys

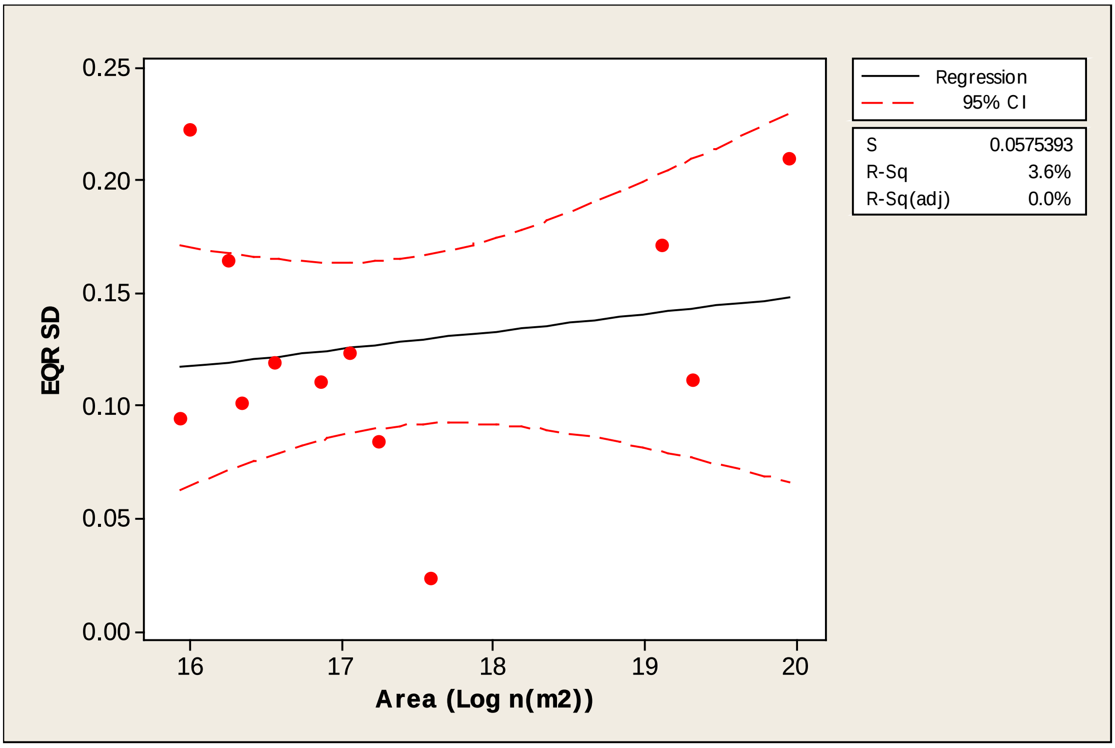


Figure 6.5 EQR standard deviation versus water body area ($\log_n(m^2)$) with upper and lower 95% confidence intervals (CIs) for transitional water subtidal WFD 2007-2009 surveys

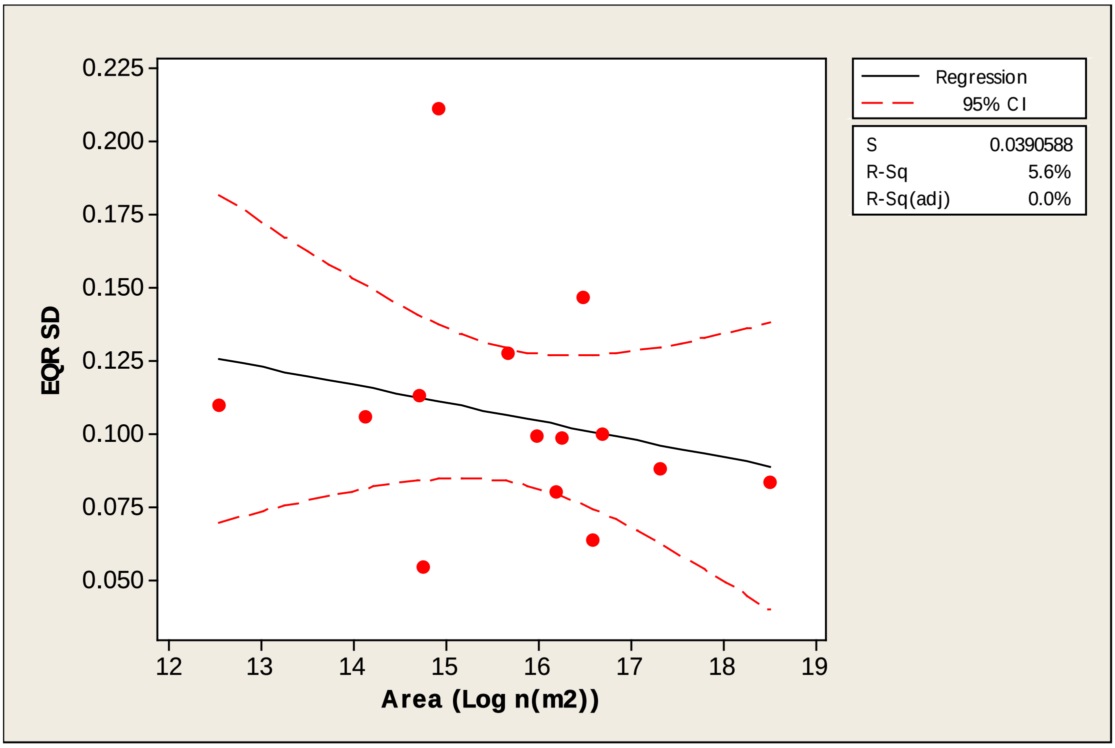


Figure 6.6 EQR standard deviation versus water body area ($\log_n(m^2)$) with upper and lower 95% confidence intervals (CIs) for transitional water intertidal WFD 2007-2009 surveys

Table 6.2 Regression analysis ANOVA results for testing effect of water body area ($\log_n(m^2)$) and EQR variability for differing coastal and transitional water monitoring conditions

Water body category	Tidal exposure	n	*F*	*p* value
Coastal water	Subtidal	13	0.01	0.935*
	Intertidal	3	Insufficient data	Insufficient data
Transitional water	Subtidal	12	0.38	0.554*
	Intertidal	14	0.71	0.414*

Note: * Not significant (*p* >0.05).
n = number of water bodies in test data
F = test statistic for comparison against threshold to determine significance

These results support the change in survey design to dissociate sample effort from water body area.

6.8 Discussion

Systematic error can only be identified and incorporated into the RoM if it has been measured, that is, the RoM cannot take into consideration systematic error in the data that is not observed, whether due to measurements of a potentially influential physicochemical parameter not being taken due to practical limitations, or from an influential factor that is not feasible to measure within the monitoring programme (for example, the impacts of predating bird populations on intertidal assemblages).

In the case of the data used in calculating reference conditions within the IQI, for example, there is potential for systematic error to exist in salinity data where spot measured salinity may be biased (recorded values higher than true values) as measurements fail to represent salinity at low tidal states where there is a greater influence of freshwater (one of the main environmental influential factors in benthic assemblage structure). This would result in bias in the reference values (overestimated taxa number, 1-(AMBI/7) and 1-λ′) with potentially underestimated EQR values as a result.

Locations vulnerable to acute short term yet high magnitude changes in physicochemical conditions (for example, short periods of extreme low salinity, dissolved oxygen, infrequent aerial exposure) are also prone to bias that may be difficult to observe if such factors are not monitored frequently enough to observe such events. This highlights the need to always consider the potential influence of local environmental conditions when interpreting IQI results.

Understanding variability is crucial at the survey planning phase to ensure the monitoring data provide the greatest statistical confidence to test the hypothesis being addressed. In addition to minimising variability through careful planning at the survey design phase (for example, exclude components of variability that are not necessary in addressing the test hypothesis), it is also critical that:

- measurement error is minimised by establishing the correct sample collection and analysis protocols
- systematic error is minimised through an understanding of how the measured parameters (for example, the metrics within the IQI) are influenced by environmental factors and methodology

The approach to separating models to estimate EQR variability according to salinity (that is, whether a coastal or transitional water) and air exposure (intertidal/subtidal) fails to acknowledge that further environmental factors may also influence EQR variability. Factors such as substratum characteristics or depth may further influence variability. For example, elevated taxa number in highly mixed sediments may serve to stabilise AMBI values in contrast to highly sorted sediments where relatively few taxa exist. The further separation of variability models according to changes in these additional environmental factors could enable EQR variability to be set according to, for example, the dominant habitat within a water body to further improve the precision of estimates of CofC/RoM and sample effort through power analysis. With sufficient data, there is the potential to estimate variability along a continuum (in a similar vein to how reference conditions are set) rather than setting them according to discrete habitat divisions. It is recommended that further research is undertaken in this area.

While additional environmental factors (for example, depth, PSA characteristics) may influence variability, these are known with less precision at the survey planning stage and are therefore not currently factored into the survey design. Only where such physicochemical conditions are known a priori during the survey planning stage could variability estimates (and therefore sample effort) be adapted according to, for example, substratum. There is the potential for such information to be included in the planning of surveys where such physicochemical data exist which may be used to improve the efficiency of monitoring programmes in terms of achieving consistency in confidence for assessment results (discussed further in Chapters 7 and 8).

7 Confidence of classification

As water body assessments are based on estimations, there is a likelihood that the ecological status of a water body as derived from the sample population will differ from the ecological status of the entire population. This chapter describes the process of using EQR variability estimates to calculate the confidence of class (CofC) and corresponding risk of misclassification (RoM) for a water body assessment.

7.1 Introduction

Providing an estimate of the statistical uncertainty of water body assessments is a statutory requirement of the WFD (Annex V, section 1.3). As described in Chapter 6, water body assessments based on estimates of ecological quality from sample data are subject to elements of variability (spatial, temporal, sampling and measurement).

When assigning discrete ecological status classes, variability means that, depending on the proximity of the water body assessment result to a class boundary, there is a likelihood that the 'true' status (that is, that status if the EQR for the total population was known with zero error) is different to that assigned. This is termed the 'risk of misclassification' (RoM). Conversely, the statistical confidence that the status assigned from the sample population falls into each of the five ecological status classes is referred to as the 'confidence of class' (CofC).

The described approach to RoM is applicable to assessments based on mean EQRs from multiple samples where values (or transformed values) are approximately normally distributed about the mean. RoM and CofC are terms that define the statistical confidence of the assessments being 'correct'. They do not provide a measure of how true an assessment is in relation to the WFD normative definitions, rather they are a measure of the water body EQR in relation to the set class boundaries.

The calculation of the RoM and CofC uses estimates of EQR variability following the approach described in Chapter 6. It is important to note that variability for a given dataset should be estimated only from data where the components of variability are comparable, that is, similar sample collection/processing methods and habitat (further investigation is needed into temporal effects on variability).

7.2 Basis for estimating uncertainty

The approach developed to define and report the CofC and RoM for WFD coastal and transitional water benthic invertebrates was developed in a collaborative project between the Environment Agency and the Water Research Centre (WRc). The approach is also described by Ellis and Adriaenssens (2006) in terms of its general application to the WFD ecological quality elements. Unlike power analysis, which is used a priori to estimate sample numbers at the survey design stage of a monitoring programme (see Chapter 8), CofC and RoM are established a posteriori using empirical EQR data.

The approach to CofC and RoM requires the following information for a given assessment:

- mean EQR

- ecological class status boundaries
- standard error of the assessment data

The approach is based on the assumption that the variability of individual sample EQRs about the mean is normally distributed. However, this assumption does not hold across the full range of values on the EQR zero to one scale: where the mean EQR is close to zero, values are restricted in terms of their departure below the mean (limited by zero) but less constrained in terms of their departure above the mean, and vice versa when the mean is close to one (see Chapter 6). To account for this and apply the assumption of normality, the CofC and RoM calculations should be based on EQR values and associated boundaries on a log scale. Log transformation results in the EQR operating between -∞ and +∞, thus eliminating the issue of spill outside the zero to one scale.

7.3 Calculating confidence of class (CofC)

Using estimated values of standard error, mean EQR and class status boundaries, the probability of the true population mean EQR occurring above and below the ecological status class assigned by the sample mean EQR can be expressed by Equations 7.1 and 7.2 respectively.

True population mean above assigned status

$$\text{Probability of value occurring above assigned status} = \text{NORMDIST}\left(\frac{\text{EQR - Upper boundary}}{\text{SE}}\right)$$

Equation 7.1

True population mean below assigned status

$$\text{Probability of value occurring below assigned status} = \text{NORMDIST}\left(\frac{\text{Lower boundary - EQR}}{\text{SE}}\right)$$

Equation 7.2

The NORMDIST function (Microsoft® Excel) calculates cumulative probabilities for the standard normal distribution (mean of zero, standard deviation of one). These values can be used to calculate the probability that the true population mean EQR falls within the ecological status class assigned by the sample mean EQR, or confidence of class using the following formula:

$$\text{Confidence of Class (\%)} = 100 \times \left(1 - \left(\begin{array}{l}\text{Probability of value occurring above assigned status} \\ + \text{Probability of value occurring below assigned status}\end{array}\right)\right)$$

Equation 7.3

The EQR standard deviation, and corresponding standard error (for a given set of replicates and so on), has to be estimated at each point over the EQR scale. Using Equation 7.3, the probability (%) that the true population status will fall within the status class as assigned by the sample EQR can be calculated for each value over the EQR scale (see example in Figure 7.1).

As a result of the EQR operating between zero and one, variability is at a minimum at either end of the EQR scale (standard error approaches zero) resulting in high CofC where EQR values near zero or one. Variability reaches a maximum near the midpoint of the EQR scale (although this may differ depending on the variability characteristics for a given set of data). At approximately the midpoint of each status class, CofC reaches its maximum and decreases as the EQR departs from this value (Figure 7.1). Depending on the degree of variability within the dataset, this may result in high degrees of overlap between CofC for EQRs of poor, moderate and good status. At the status boundaries, the likelihood of the true population EQR falling either side of the boundary is equal, so CofC has a maximum of 0.5. CofC on the boundaries may fall below 0.5 where there is a probability that the true population mean falls within greater than two status classes.

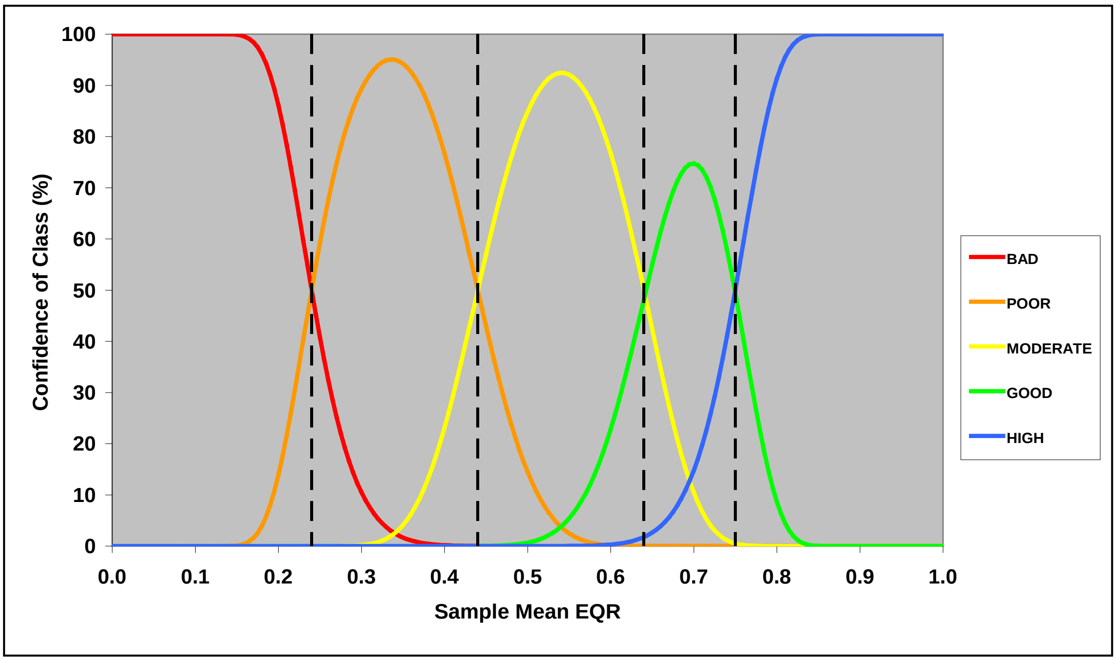


Figure 7.1 Confidence (%) that the true population EQR falls within each of the five ecological status classes (CofC) over the EQR scale

Note: Standard error based on five samples using standard deviation estimates from data from Environment Agency WFD 2007-2009 surveillance monitoring data from subtidal transitional waters (0.1 m^2 grab, 0.5 mm sieve mesh)

The probabilities of the true population EQR falling below, inside and above each ecological status as assigned by sample EQR are given in Table 7.1.

Table 7.1 Probabilities of the true population EQR falling below, inside and above each ecological status as assigned by sample EQR, including estimated SD and associated SE

Mean EQR	Estimated SD	No. of samples	Estimated SE	BAD			POOR			MODERATE			GOOD			HIGH		
				Below	Inside	Above	Below	Inside	Above	Below	Inside	Above	Below	Inside	Above	Below	Inside	Above
0.00	0.001	5	0.000	0.00	**1.00**	0.00	1.00	**0.00**	0.00	1.00	**0.00**	0.00	1.00	**0.00**	0.00	1.00	**0.00**	0.00
0.05	0.025	5	0.011	0.00	**1.00**	0.00	1.00	**0.00**	0.00	1.00	**0.00**	0.00	1.00	**0.00**	0.00	1.00	**0.00**	0.00
0.10	0.047	5	0.021	0.00	**1.00**	0.00	1.00	**0.00**	0.00	1.00	**0.00**	0.00	1.00	**0.00**	0.00	1.00	**0.00**	0.00
0.15	0.066	5	0.030	0.00	**1.00**	0.00	1.00	**0.00**	0.00	1.00	**0.00**	0.00	1.00	**0.00**	0.00	1.00	**0.00**	0.00
0.20	0.082	5	0.037	0.00	**0.86**	0.14	0.86	**0.14**	0.00	1.00	**0.00**	0.00	1.00	**0.00**	0.00	1.00	**0.00**	0.00
0.25	0.096	5	0.043	0.00	**0.41**	0.59	0.41	**0.59**	0.00	1.00	**0.00**	0.00	1.00	**0.00**	0.00	1.00	**0.00**	0.00
0.30	0.107	5	0.048	0.00	**0.11**	0.89	0.11	**0.89**	0.00	1.00	**0.00**	0.00	1.00	**0.00**	0.00	1.00	**0.00**	0.00
0.35	0.116	5	0.052	0.00	**0.02**	0.98	0.02	**0.94**	0.04	0.96	**0.04**	0.00	1.00	**0.00**	0.00	1.00	**0.00**	0.00
0.40	0.122	5	0.054	0.00	**0.00**	1.00	0.00	**0.77**	0.23	0.77	**0.23**	0.00	1.00	**0.00**	0.00	1.00	**0.00**	0.00
0.45	0.125	5	0.056	0.00	**0.00**	1.00	0.00	**0.43**	0.57	0.43	**0.57**	0.00	1.00	**0.00**	0.00	1.00	**0.00**	0.00
0.50	0.127	5	0.057	0.00	**0.00**	1.00	0.00	**0.14**	0.86	0.14	**0.85**	0.01	0.99	**0.01**	0.00	1.00	**0.00**	0.00
0.55	0.125	5	0.056	0.00	**0.00**	1.00	0.00	**0.02**	0.98	0.02	**0.92**	0.05	0.95	**0.05**	0.00	1.00	**0.00**	0.00
0.60	0.122	5	0.054	0.00	**0.00**	1.00	0.00	**0.00**	1.00	0.00	**0.77**	0.23	0.77	**0.23**	0.00	1.00	**0.00**	0.00
0.65	0.116	5	0.052	0.00	**0.00**	1.00	0.00	**0.00**	1.00	0.00	**0.42**	0.58	0.42	**0.55**	0.03	0.97	**0.03**	0.00
0.70	0.107	5	0.048	0.00	**0.00**	1.00	0.00	**0.00**	1.00	0.00	**0.10**	0.90	0.10	**0.75**	0.15	0.85	**0.15**	0.00
0.75	0.096	5	0.043	0.00	**0.00**	1.00	0.00	**0.00**	1.00	0.00	**0.01**	0.99	0.01	**0.49**	0.50	0.50	**0.50**	0.00
0.80	0.082	5	0.037	0.00	**0.00**	1.00	0.00	**0.00**	1.00	0.00	**0.00**	1.00	0.00	**0.09**	0.91	0.09	**0.91**	0.00
0.85	0.066	5	0.029	0.00	**0.00**	1.00	0.00	**0.00**	1.00	0.00	**0.00**	1.00	0.00	**0.00**	1.00	0.00	**1.00**	0.00
0.90	0.047	5	0.021	0.00	**0.00**	1.00	0.00	**0.00**	1.00	0.00	**0.00**	1.00	0.00	**0.00**	1.00	0.00	**1.00**	0.00
0.95	0.025	5	0.011	0.00	**0.00**	1.00	0.00	**0.00**	1.00	0.00	**0.00**	1.00	0.00	**0.00**	1.00	0.00	**1.00**	0.00
1.00	0.001	5	0.000	0.00	**0.00**	1.00	0.00	**0.00**	1.00	0.00	**0.00**	1.00	0.00	**0.00**	1.00	0.00	**1.00**	0.00

Note: SD estimates based on data from Environment Agency WFD 2007-2009 surveillance monitoring data from subtidal transitional waters (0.1 m^2 grab, 0.5 mm sieve mesh).

7.4 Confidence of less than good status

The approach to confidence of class can be applied to water body classification in relation to the moderate/good boundary. The significance of the moderate/good boundary is that failing to achieve good ecological status will require the implementation of programmes of measures.

The probability of the true population mean EQR occurring above and below the moderate/good (M/G) boundary can be expressed using Equations 7.4 and 7.5 respectively.

True population mean above moderate–good boundary

$$\text{Probability of value occurring above M/G boundary} = \text{NORMDIST}\left(\frac{\text{EQR - M/G boundary}}{\text{SE}}\right)$$

Equation 7.4

True population mean below moderate–good boundary

$$\text{Probability of value occurring below M/G boundary} = \text{NORMDIST}\left(\frac{\text{M/G boundary - EQR}}{\text{SE}}\right)$$

Equation 7.5

Similar to the approach described for CofC (Section 7.3), these values can be used to calculate the probability that the true population mean EQR falls below the moderate/good boundary as assigned by the sample mean EQR, or confidence of failure (%) using the following formula:

$$\text{Confidence of Failure (\%)} = 100 \times \left(\text{Probability of value occurring below M/G boundary}\right)$$

Equation 7.6

As with CofC, each point over the EQR scale has an associated estimated probability that the true population EQR will fall below the moderate/good boundary as assigned by the sample EQR (Table 7.2).

Table 7.2 Probabilities of the true population EQR falling below and above the moderate/good (M/G) boundary as assigned by sample EQR, estimated SD, associated SE, and confidence of failure (%) of achieving greater than or equal to good ecological status

Mean EQR	Estimated SD	No. of samples	Estimated SE	Status	Below M/G	Above M/G	Confidence of failure (%)
0.00	0.001	5	0.000	Bad	1.00	0.00	100.0
0.05	0.025	5	0.011	Bad	1.00	0.00	100.0
0.10	0.047	5	0.021	Bad	1.00	0.00	100.0
0.15	0.066	5	0.030	Bad	1.00	0.00	100.0
0.20	0.082	5	0.037	Bad	1.00	0.00	100.0
0.25	0.096	5	0.043	Poor	1.00	0.00	100.0
0.30	0.107	5	0.048	Poor	1.00	0.00	100.0
0.35	0.116	5	0.052	Poor	1.00	0.00	100.0
0.40	0.122	5	0.054	Poor	1.00	0.00	100.0
0.45	0.125	5	0.056	Moderate	1.00	0.00	100.0
0.50	0.127	5	0.057	Moderate	0.99	0.01	99.3
0.55	0.125	5	0.056	Moderate	0.95	0.05	94.6
0.60	0.122	5	0.054	Moderate	0.77	0.23	76.9
0.65	0.116	5	0.052	Good	0.42	0.58	42.3
0.70	0.107	5	0.048	Good	0.10	0.90	10.5
0.75	0.096	5	0.043	High	0.01	0.99	0.5
0.80	0.082	5	0.037	High	0.00	1.00	0.0
0.85	0.066	5	0.029	High	0.00	1.00	0.0
0.90	0.047	5	0.021	High	0.00	1.00	0.0
0.95	0.025	5	0.011	High	0.00	1.00	0.0
1.00	0.001	5	0.000	High	0.00	1.00	0.0

Note: SD estimates based on data from Environment Agency WFD 2007-2009 surveillance monitoring data from subtidal transitional waters (0.1 m^2 grab, 0.5 mm sieve mesh)

7.5 Calculating risk of misclassification (RoM)

The RoM (%) for a given EQR is calculated as:

$$\text{Risk of Misclassification (\%)} = 100 - \text{Confidence of Class (\%)} \qquad \text{Equation 7.7}$$

The data displayed in Table 7.1 can be used to calculate the overall CofC and associated RoM over the full EQR scale (Table 7.3).

Table 7.3 Estimated SD over the EQR scale with associated SE (based on five samples), ecological status, CofC and RoM

Mean EQR	Estimated SD	No. of samples	Estimated SE	Ecological status	CofC (%)	RoM (%)
0.00	0.001	5	0.000	Bad	100.0	0.0
0.05	0.025	5	0.011	Bad	100.0	0.0
0.10	0.047	5	0.021	Bad	100.0	0.0
0.15	0.066	5	0.030	Bad	99.9	0.1
0.20	0.082	5	0.037	Bad	86.1	13.9
0.25	0.096	5	0.043	Poor	59.2	40.8
0.30	0.107	5	0.048	Poor	89.3	10.7
0.35	0.116	5	0.052	Poor	94.2	5.8
0.40	0.122	5	0.054	Poor	76.7	23.3
0.45	0.125	5	0.056	Moderate	57.0	43.0
0.50	0.127	5	0.057	Moderate	84.8	15.2
0.55	0.125	5	0.056	Moderate	92.1	7.9
0.60	0.122	5	0.054	Moderate	76.7	23.3
0.65	0.116	5	0.052	Good	55.0	45.0
0.70	0.107	5	0.048	Good	74.7	25.3
0.75	0.096	5	0.043	High	50.0	50.0
0.80	0.082	5	0.037	High	91.3	8.7
0.85	0.066	5	0.029	High	100.0	0.0
0.90	0.047	5	0.021	High	100.0	0.0
0.95	0.025	5	0.011	High	100.0	0.0
1.00	0.001	5	0.000	High	100.0	0.0

Note: Variability estimated from 0.1 m^2 grab transitional water data processed using a 0.5 mm sieve.

Plotting the RoM across the EQR range produces results in a ‘washing-line’ pattern (Figure 7.2), illustrating the characteristics of the RoM as lowest at the centre of each status class and highest at each boundary as described previously. As illustrated and given the standard error associated with the data, the RoM does not reach zero between poor and good ecological status, and reaches a minimum RoM for good ecological status of approximately 25% which is influenced by the relatively narrow width of the status. The effect of reducing overall standard error as a consequence of sample effort can be observed by plotting the RoM over the EQR scale for assessments based on 3 and 15 samples (Figure 7.3).

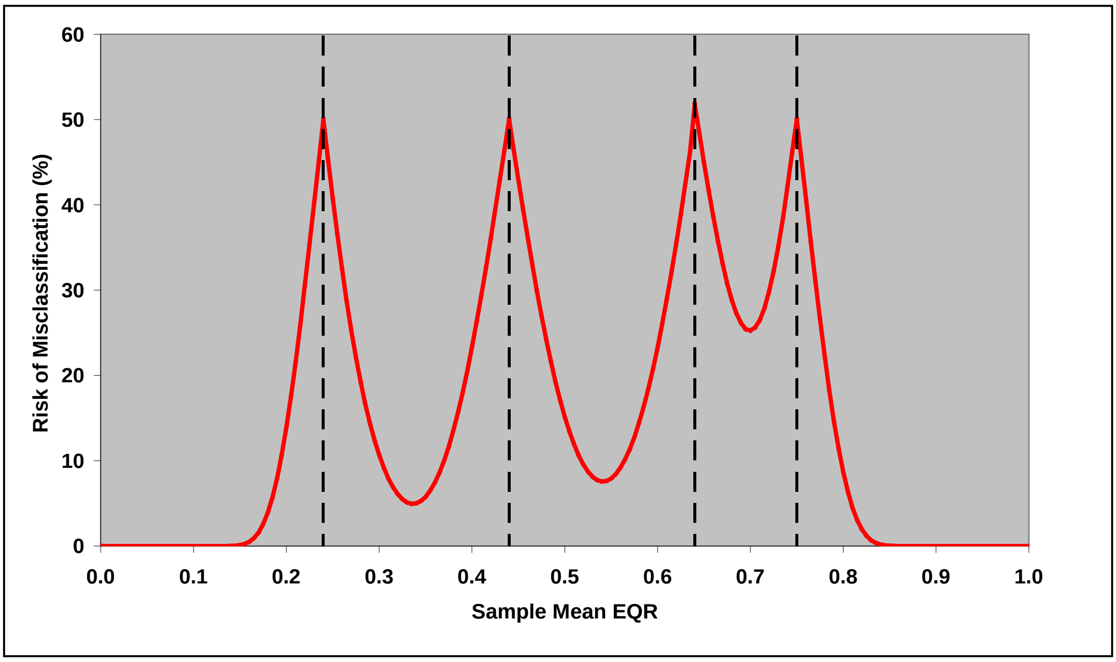


Figure 7.2 Risk of misclassification over the EQR scale (based on five samples)

Note: Variability estimated from 0.1 m^2 grab transitional water data processed using a 0.5 mm sieve.

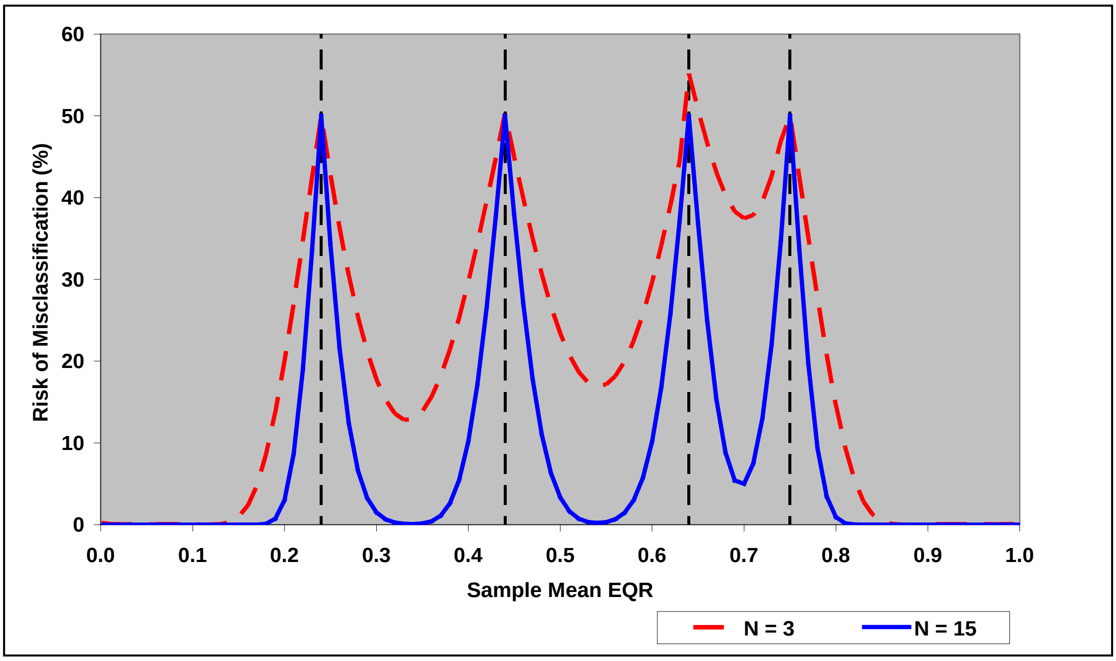


Figure 7.3 Risk of misclassification over the EQR scale based on standard error values calculated on the basis of 3 and 15 samples

Note: Variability estimated from 0.1 m^2 grab transitional water data processed using a 0.5 mm sieve.

Increasing sample numbers generally reduces the RoM. In the illustrations, increasing effort from 3 to 15 samples reduces the RoM to close to zero near the centre of poor and moderate ecological status, and from ~37% to ~5% near the centre of 'good' status. In all cases, however, the increase in sample effort will not reduce the RoM at the status boundaries to below 50% for the reasons highlighted previously.

The implications on the RoM of sampling in conditions with differing degrees of natural variability can be observed by comparing RoM versus EQR plots for coastal (low relative variability) and transitional (higher relative variability) data from subtidal muds (Figure 7.4). It should be noted that the transitional water data on which the illustration below is based come from WFD data with a narrow range of salinity (~20 to ~30). The risk of misclassification in transitional waters is anticipated to increase if samples were collected over the full range of salinity.

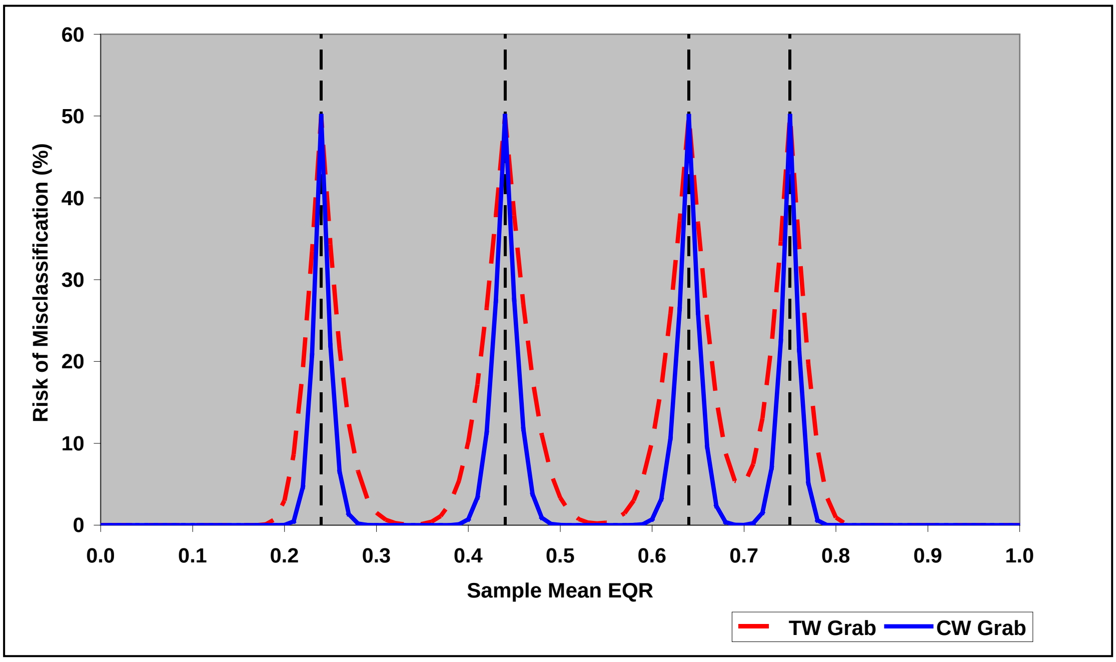


Figure 7.4 Risk of misclassification over the EQR scale based on standard error values calculated on the basis of 15 samples from subtidal, muddy sediments in transitional (TW) and coastal (CW) waters

Note: Variability estimated from 0.1 m^2 grab and 0.5 mm transitional water/1 mm coastal water data

In the illustration, the methodology (sieve mesh aperture) also differs between transitional and coastal water data which may contribute to the variability.

Preliminary analysis of CSEMP data (1999-2006) was undertaken to identify whether EQR variability differs according to sieve mesh size. The data consisted of 179 macrobenthic samples (including 0.1 m^2 Day grab and ~0.01 m^2 intertidal core methods) processed with 0.5 and 1 mm mesh sieves. Each CSEMP survey consists of more than one replicate taken at a single site on a single sampling occasion. For the two sample collection methods, ANOVA of within-survey standard deviation (that is, within-site and measurement variability) was undertaken to identify whether EQR variability was affected by sieve mesh size (Table 7.4).

Table 7.4 Variability (within-site and measurement) for 0.5 and 1 mm data for CSEMP data

Method	No. of samples [1]	Mean SD (0.5 mm)	Mean SD (1 mm)	SD difference	p [2]
Day grab	151	0.040	0.045	0.005	0.212*
Core	28	0.051	0.080	0.029	0.008

Notes: [1] 179 samples taken between 1999 and 2006
[2] Probability of difference between 0.5 mm and 1 mm mean SD values
* Not significant (p >0.05)

While variability is generally lower for 0.5 mm data relative to 1 mm data for both sample collection methodologies, the standard deviation is only significant at 95% confidence for the core data. The negligible difference between the EQR standard deviations for 0.5 mm and 1 mm data suggests that the elevated variability observed in similar substratum in transitional waters compared with coastal waters is due to the effect of increasingly variable physicochemical conditions (including measurement error) rather than differences in sieve mesh size.

7.6 Environment Agency's WFD surveillance monitoring

The Environment Agency's current approach to WFD surveillance monitoring bases sampling effort on the variability in accordance with water body category (coastal/transitional) and sampling method (subtidal grab/intertidal core). The current approach to estimating variability only divides data according to these categories. The approach does not factor in the potential influence of sediment type on variability, as the WFD data used in developing the models was based on a narrow range of sediment types.

The adaptation and updating of reference conditions to enable a broader range of habitats to be assessed by the IQI will allow between-substratum differences in variability to be established with the potential to further set sample effort according to within-water body conditions. Chapter 8 describes how EQR variability is incorporated within power analysis to derive sampling effort to achieve predetermined degrees of confidence in assessment results.

Adapting sampling effort according to water body category (and potentially habitat type) and method specific variability allows CofC and RoM to be kept fairly constant. Conversely, setting sample effort independent of variability results in inconsistencies in CofC and RoM. The implications to the RoM as a result of fixing sample effort across all types and methods, and adapting sample effort according to variability are illustrated in Figures 7.5 and 7.6 respectively.

As described in Chapter 6, variability estimates and corresponding uncertainty estimates are dependent on the components of variation to which the data are subjected. It should not be assumed that the values for CofC or RoM can be applied to EQRs from all survey data; they may be applied if the components and associated magnitude of variation are comparable.

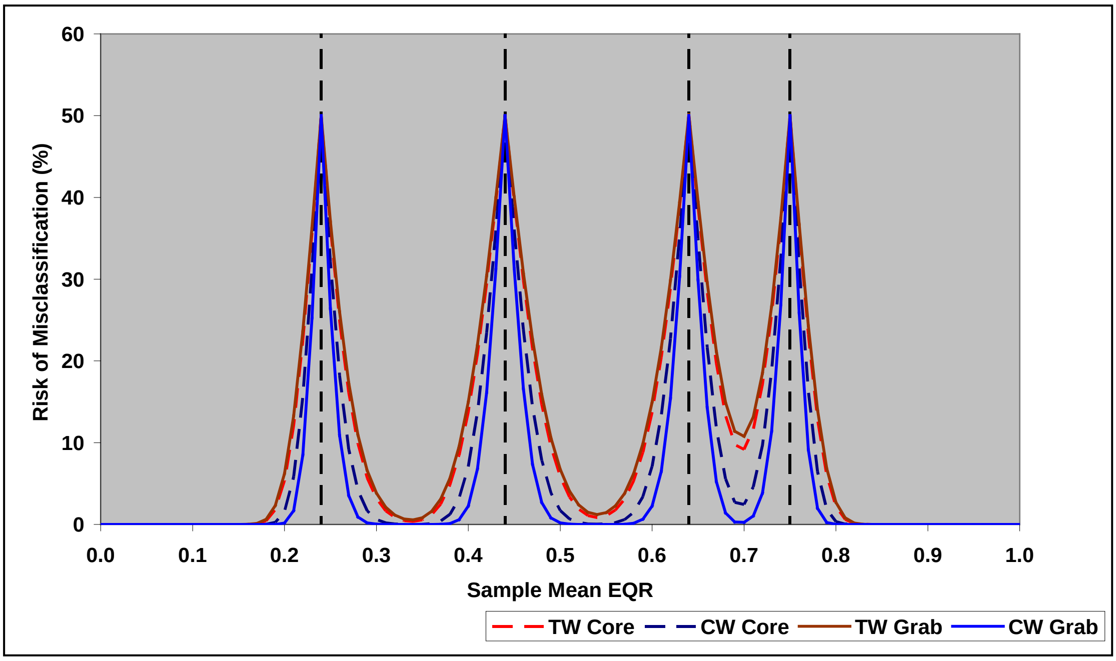


Figure 7.5 RoM over the EQR scale for different water body categories and sample collection methods based on fixed sample effort

Notes: Variability estimated from Environment Agency WFD 2007-2009 surveillance monitoring data.
TW = transitional water, CW = coastal water
n = 10

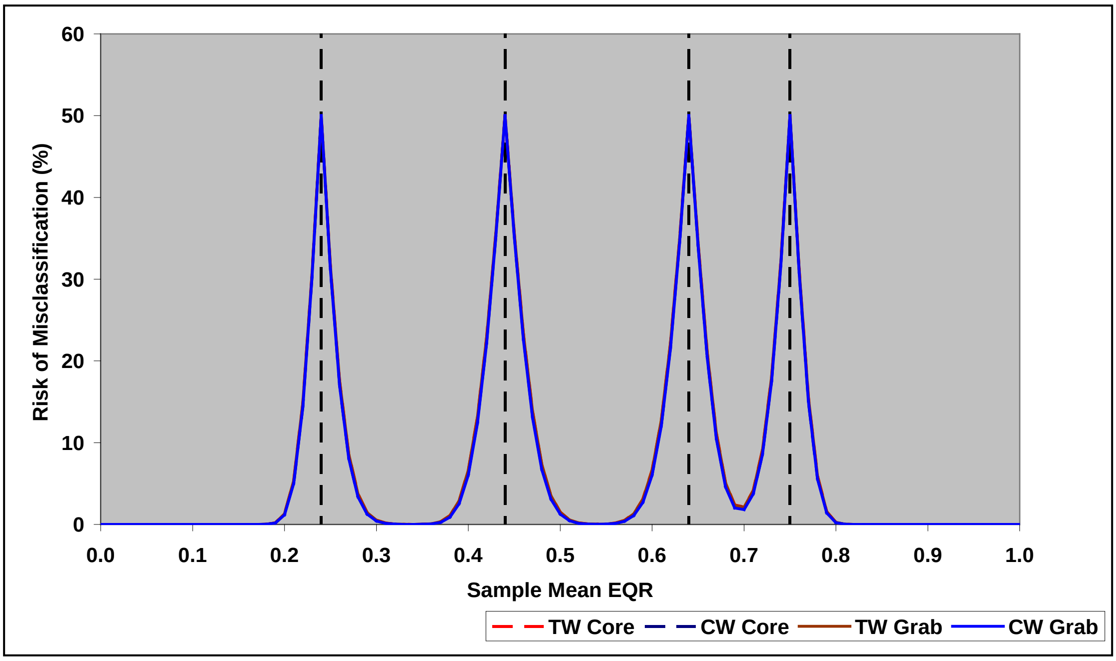


Figure 7.6 RoM over the EQR scale for different water body categories and sample collection methods based on water body category and sample collection method specific sample effort

Notes: Variability estimated from Environment Agency WFD 2007-2009 surveillance monitoring data.
TW = transitional water, CW = coastal water
TW Core n = 19, CW Core n = 11, TW Grab n = 21, CW Grab n = 6
As sample effort has been adapted to achieve constant standard error values between water body categories and sample collection methods, RoM curves become almost exactly superimposed.

7.7 Discussion

Risk of misclassification relates only to the statistical confidence in the sample mean EQR values in contrast to the expected population mean EQR. RoM is not intended to be a representation of the effectiveness or sensitivity of the classification tools in response to pressures; this has to be addressed in EQR pressure/response studies. Equally, RoM is not intended to act as an expression of how appropriately the interpretation of the requirements of the WFD have been expressed within a classification tool (that is, whether a status as assigned by a classification tool correctly conforms to the requirements of the normative definitions). This is addressed through the boundary setting process.

The approach to the risk of misclassification does not currently factor in the variability of the estimated reference conditions. Error arises in measuring the environmental variables at a site and so there is some statistical uncertainty in the 'expected' value (for example, number of taxa). Thus there is error due to both the 'observed' and 'expected' components of the EQR.

This aspect of the data is shared by the RIVPACS III+ (River Invertebrate Prediction and Classification System) system for the classification of riverine macroinvertebrate fauna (Wright et al. 2000). RIVPACS III+ uses biological and physicochemical data from pristine freshwater sites to predict the invertebrate structure from a set of physicochemical conditions and adopts a simulation analysis approach to apply a degree of confidence to the reference conditions.

Due to IQI reference conditions being estimated from numerical models (Chapter 4), it would be incorrect to assume that they are without error, and comparable means for assigning confidence is required. As the reference models are based on regression formulae, it may be possible to establish confidence intervals to apply to the derived reference condition values to incorporate into the variability estimates. Without factoring this additional level of uncertainty into the classifications, it is likely that current RoM values are underestimates.

The current model provides an initial approach of how to calculate RoM. Ultimately, with sufficient data, a RoM model may potentially be derived for each habitat–salinity combination as appropriate within both the intertidal and subtidal.

8 Use of power analysis for WFD monitoring

The statistical confidence behind an assessment of a water body depends on the variability and quantity of the underlying sample data. To ensure that the WFD monitoring programme data provide the required statistical confidence in the subsequent assessment, sample effort must be derived accordingly. This chapter describes the process adopted to calculate the sampling effort required to detect a difference between an EQR and a given threshold for a prescribed degree of confidence.

8.1 Introduction

Power analysis is a method adopted to estimate the number of samples required to identify whether the results of a test are sufficient to draw the correct conclusions regarding the state of the real world, that is, what is the likelihood that a null hypothesis (H_0) is rejected when it should be, and not rejected when it should not be. It is an important stage in the experimental design process, allowing the determination of the sample effort required to detect an effect of a given size (for example, a 0.1 difference in EQR values) for a specified degree of confidence.

Power analysis is based on four main interacting components:

- sample number
- the size of the effect
- the degree of significance/confidence (Type I error)
- the power of finding the effect (Type II error)

Any three components can be used to calculate the fourth.

Resources and technical feasibility rarely allow for limitless sampling effort in environmental monitoring programmes. When planning such programmes, power analysis enables efficient use of resources, ensuring that evidence provided by the monitoring programme is sufficient for management action (that is, the required probability of detecting an effect of a given size, with a given level of confidence is achieved). National monitoring programmes often require monitoring to be undertaken at sites exposed to differing degrees of variability (for example, transitional versus coastal waters). Power analysis enables resources (for example, sample numbers) to be modified for the varying circumstances, ensuring sufficient statistical confidence while preventing excessive sampling. Power analysis may also provide the justification to abandon a programme if the available resources do not allow or it is not technically feasible to obtain sufficient data to attain an acceptable degree of confidence in the results.

Risk of misclassification provides an *a posteriori* estimation of the uncertainty in an assessment result by relating average EQR values to WFD status boundaries once data is collected (Chapter 7). Power analysis is an a priori method to estimate the spatial and temporal configuration of samples required to ensure that the assessment results will provide a predetermined level of precision to an accepted level of confidence (expressed as a probability).

WFD requires monitoring from a range of habitats within coastal and transitional water body types. As noted in Chapter 6, EQR variability differs according to water body category and sample methodology. Applying a fixed sampling effort for all surveys would result in varying degrees of statistical confidence. By incorporating water body category and method specific variability into power analysis, sample numbers can be adapted accordingly to standardise the degree of precision in the assessment results across all water bodies.

The approach to power analysis follows the recommendations made in 2006 by Paul Somerfield and Bob Clarke of the Plymouth Marine Laboratories.

8.2 Type I and II error

Rejection of the null hypothesis from a test is often considered statistically significant where $p < 0.05$. However, in some cases, the null hypothesis should be rejected but the test may not have been powerful enough to detect it as, for example, $p < 0.05$. To conclude that the null hypothesis should be rejected (for example, a statistical difference exists between a water body EQR and an ecological class boundary), it must be demonstrated that the test had sufficient power to detect it.

Analysis of variance (ANOVA) is subject to Type I (α) and Type II (β) error. Type I error is the probability that a null hypothesis (H_0) is rejected but where it is actually true (the probability that we have found something that actually does not exist). Type II error is the probability of failing to detect a difference in mean response but where a difference does actually exist that is, failing to reject H_0 when H_0 is false in reality (Table 8.1). Type II error can be termed the power of a statistical test and is calculated as 1-Power. Power measures the likelihood of a test in reaching the correct conclusion, that is, accepting H_0 when it should be accepted and rejecting H_0 when it should be rejected.

Table 8.1 Types of error in hypothesis testing

	H_0 is true	**H_0 is false**
H_0 is rejected	Type I error	No error
H_0 is not rejected	No error	Type II error

In formulating a monitoring programme, the resulting data should be powerful enough to detect a difference between treatments, or in the case of WFD, a difference between an observed EQR and a class boundary on the EQR scale.

The number and spatial/temporal configuration of samples within a test (or survey) is crucial to ensure that the results hold enough power to draw the correct conclusion, and accept or reject H_0 accordingly. Without demonstrating that a test is powerful enough to detect a difference between treatments (or EQR and ecological status boundary), conclusions cannot be drawn as to whether a real difference exists and whether management action is required.

8.3 Approach to power analysis

To estimate the number of samples required to detect a difference in mean EQR response in the IQI relative to a given class status boundary, the following information is needed:

- size of effect (difference in EQRs to be detected) (δ)
- residual variance (error) of single EQRs (σ)
- normal distribution function (Φ)
- Type I error (probability of rejecting H_0 when H_1 should be rejected) (p)
- 1-Type II error (probability of failing to reject H_0 when it should be rejected) (P)

The approximate number of samples (n) required to identify whether the test can detect δ for predetermined values of P first requires the calculation of kappa or κ (the *p*th moment about the mean (Zar 1984)) as follows:

$$\kappa = \left(\frac{\sigma}{\delta}\right)^2 \times \left[2 + \Phi^{-1}(P)\right]^2 \qquad \text{Equation 8.1}$$

For example, if a survey is planned to provide 90% confidence (P) in detecting a difference (δ) in EQR of 0.1 from the moderate/good status boundary where the estimated standard deviation (σ) is 0.08, substituting the following values into the formula:

σ = 0.08

δ = 0.1

P = 0.9

calculates κ as:

$$\kappa = \left(\frac{0.08}{0.1}\right)^2 \times \left[2 + \Phi^{-1}(0.9)\right]^2 = 6.89 \qquad \text{Equation 8.2}$$

The estimated number of samples (n) is calculated from κ as follows:

$$n > (\kappa + 1) + \sqrt{(\kappa^2 + 1)} \qquad \text{Equation 8.3}$$

Using the above example where κ = 6.89, the number of samples (n) can be determined as:

$$n > (6.89 + 1) + \sqrt{(6.89^2 + 1)} => 14.86 \qquad \text{Equation 8.4}$$

This equates to a required sample number (n) of 15 when rounded to the nearest whole sample.

Applying the approach to a range of scenarios (in terms of differing levels of confidence and detectable size of effect) can provide a valuable tool in assisting the planning of a monitoring programme given resource limitations (budget, time, personnel availability and so on).

8.4 Environment Agency’s WFD surveillance monitoring programme

The IQI results obtained from the Environment Agency’s initial WFD surveillance monitoring (2007-2009) were analysed to validate, and where necessary revise, the required sample effort. The approach to estimating water body and sampling method specific variability for the surveillance monitoring data is described in Chapter 6.

The required parameters for power analysis (as highlighted in Section 8.3) were proposed by the Environment Agency in 2009[11] for the second stage of WFD surveillance monitoring (2010-2012) as follows:

- size of effect (δ) = 0.1
- Type II error/required confidence (P) = 0.9 (that is, 90%)

EQR values are proportions. As a result, the left-hand and right-hand tails of the distribution are truncated by theoretical minimum and maximum of zero and one respectively, and the variance of the sample data over the EQR scale will therefore be asymmetrical or non-normal. As the approach to power analysis requires the variance of the samples about the mean to be approximately normal, values on the EQR scale therefore require arcsine transformation prior to analysis.

SD (σ) values were estimated using the variability models (Chapter 6) for each combination of water body category and sampling method. As the value for σ varies according to EQR, the EQR value upon which σ is based was defined as being the size of effect subtracted from the moderate/good boundary. In other terms, the monitoring design required that the survey data were able to state that EQR values ≥0.1 above the moderate/good boundary (EQR 0.64) achieved GES with ≥90% confidence. Therefore, the EQR value used to estimate σ was calculated as 0.64 + δ = 0.74. The standard deviation values applied to power analysis were calculated from the variability models (Table 8.2).

Table 8.2 Standard deviation values estimated from variability models for each water body category and sampling methodology to be applied to power analysis (predominantly sand/mud habitats)

Water body category	Sampling methodology	Estimated σ at EQR = 0.74
Coastal water	Subtidal grab	0.051
	Intertidal core	0.070
Transitional water	Subtidal grab	0.098
	Intertidal core	0.094

The estimated standard deviation values could then be applied to the power analysis formula with values for δ = 0.1 and P = 0.9 to estimate the sample effort required for assessment (Table 8.3).

[11] Parameters may be revised for future stages of WFD surveillance monitoring.

Table 8.3 Sample effort established for the Environment Agency WFD surveillance monitoring programme for each water body category and sampling methodology

Water body category	Sampling methodology	Sample effort
Coastal water	Subtidal grab	5.7
	Intertidal core	10.5
Transitional water	Subtidal grab	20.8
	Intertidal core	18.9

At the time of development, variability models were based on limited data (coastal water subtidal = 14, coastal water intertidal = 3, transitional water subtidal = 14, transitional water intertidal = 16), with an anticipated degree of variability in standard deviation values estimated from the models. Values established through power analysis were therefore rounded up to accommodate a potential increase in variance values as the variability models are refined as additional data becomes available. As additional data are included on which variability estimates are based, this is expected to improve the degree of confidence in the estimates and so the need to round up sample numbers as a precaution will reduce.

The WFD surveillance monitoring programmes implemented by SEPA and NIEA may differ to the programme adopted by the Environment Agency. As a consequence of differing survey designs, different components of variability may apply (for example, within-station, temporal), with implications for the overall variability of the resulting assessment data. As such, the sample effort calculated for the Environment Agency monitoring design is unlikely to have the same power to detect change if applied to the survey design of SEPA and NIEA. If EQR variability can be estimated for different survey designs, the same approach to power analysis as described in this chapter can be applied.

8.5 Discussion

It is important to stress that the aim of power analysis is to provide a method for achieving a given level of statistical precision of the EQR values for an assessment. It does not relate to the level of accuracy of the results in determining effectiveness of the classification tool in fulfilling the requirements of the normative definitions.

As power analysis is required a priori to an experiment/survey during the design stage, this has potential implications in designing new programmes (or adapting those in existence) whereby little data exist upon which to estimate the 'new' components of variance. The recommendation is to undertake a baseline survey to estimate the components of variance prior to the test survey. However, in cases where this is not possible (as a result of time or resource constraints), to improve the likelihood that sufficient samples are taken so that the survey results (once available) yield sufficient confidence, an approach may be to increase the number of samples (for example, by 10%) above that calculated by power analysis to provide a margin of error should the variance calculated a priori be an underestimate of that once the results are available.

In terms of estimating habitat-specific variability, the approach has to date only divided habitats at a broad scale on the basis of a water body category (coastal or transitional) as monitoring is targeted to equivalent sediment type in both water categories. If applied to further habitat-based subdivisions, sample effort could be further refined to improve resource allocation. For example, transitional water bodies experiencing relatively low freshwater input (and higher average salinity accordingly) may require lower sampling effort compared with those experiencing high freshwater inputs. Equally, if dependent upon substratum, if sufficient information of the substratum of a water body was known, sample effort could be further refined on the basis of more accurate variability estimates.

To estimate sample effort through power analysis for WFD, the factors that influence variability need to be known beforehand. In the case of variability linked to salinity, the designation of water bodies as either coastal or transitional provides such predetermined information, enabling the a priori calculation of sample effort accordingly. For other governing environmental factors (most notably substratum characteristics), this is to some degree unknown prior to the samples being collected. Even where substrata data are available, they are often subject to inaccuracy on the basis of the data being outdated as a result of mobile sediments, subjective due to descriptions being qualitative and classified according to differing systems or inaccurate on the basis of habitat area maps being extrapolated from sediment spot samples.

By combining methods for reliable habitat mapping with an improved understanding of how environmental factors such as substrata influence EQR variability, the power analysis process may be used to further refine sample effort to improve the efficiency of monitoring programmes. This report recommends that further research is undertaken in this area.

While the approach to power analysis may apply to similar biological metrics, it must be emphasised that the power of n samples to detect change is specific to the variance of a metric under a given set of conditions (components and magnitude of variability) and is likely to differ to that of other metrics (for example, Shannon–Weiner). The relevant specific variance of a metric must be applied to power analysis.

9 Summary and recommendations

9.1 Summary

- $IQI_{v.IV}$ was developed as a WFD compliant assessment tool to enable classification of the benthic invertebrate quality element according to the WFD normative definitions.

- Formulation of $IQI_{v.IV}$ involved the use of pressure gradient data to develop a multimetric that combines measures of taxon richness, sensitivity (AMBI) and evenness (Simpson's) of an observed benthic assemblage relative to expected values at habitat specific reference conditions to derive EQRs:

$$IQI_{\text{v.IV}} = \left(\left(0.38 \times \left(\frac{1-(AMBI/7)}{1-(AMBI_{\text{Ref}}/7)} \right) \right) + \left(0.08 \times \left(\frac{1-\lambda'}{1-\lambda'_{\text{Ref}}} \right) \right) + \left(0.54 \times \left(\frac{S}{S_{\text{Ref}}} \right)^{0.1} \right) - 0.4 \right) \Big/ 0.6$$

- $IQI_{v.IV}$ was used in the first round of WFD classifications in the UK River Basin Management Plans.

- Benthic abundance data standardisation rules to ensure consistency when analysing data from different sources (potential discrepancies in taxonomic identification and enumeration) were initially developed by in the interim technical report for the project (Prior et al. 2004). These rules were adapted in the present project (2008) to allow for the application of $IQI_{v.IV}$ to a broader range of habitats and to consider developments in benthic invertebrate analysis quality assurance protocols through NMBAQC.

- Reference conditions for coastal subtidal sand/mud habitats (0.1 m^2 grab, 1 mm sieve) were established based on the analysis of data from minimally impacted locations and expert judgement of representatives of the Marine Benthic Invertebrate Task Team (MBITT). Relationships between environmental conditions and the $IQI_{v.IV}$ metrics were identified to provide numerical models to adapt the reference conditions according to continuous environmental gradients. This enables reference conditions to be applied to a broader range of natural environmental conditions and sample collection and processing methods.

- Class status boundaries (high, good, moderate, poor and bad) for the IQI were established to comply with the WFD normative definitions on the basis of changes in benthic infaunal sensitivity (AMBI ecological group proportions) over an anthropogenic pressure gradient. Nationally proposed boundaries were adapted to maximise the agreement between the status assessments of the IQI and classification methods of the Member States of the North East Atlantic Geographical Intercalibration Group through Phase I intercalibration. The boundaries for coastal water types were formally agreed by the European Commission in 2008.

- Collaborative work between MBITT and WRc derived a method for estimating EQR variability for IQI assessments. This method was applied to enable variability to be estimated for different environmental conditions and sample collection and processing methods (coastal or transitional and intertidal or subtidal). The collaborative work also highlighted an approach to using EQR variability estimates to calculate the statistical confidence of classification (confidence of class and risk of misclassification).

- Collaborative work between MBITT and Plymouth Marine Laboratory highlighted a suitable approach to power analysis to enable EQR variability estimates to be used to determine the required sample effort to detect a given change in EQR (effect size) for a given degree of confidence. The method described was used to establish the sampling effort for Environment Agency WFD surveillance monitoring.

9.2 Recommendations

9.2.1 Data requirements of the IQI

This study has highlighted how measures of ecological health used to detect the response of an assemblage to anthropogenic pressure can be influenced by natural environmental conditions. For the IQI to be used effectively for WFD assessments, it is recommended that benthic infaunal surveys are supported by standardised environmental data (salinity and particle size analysis).

9.2.2 Maintenance of taxon lists

It is recommended that the taxon lists used for the delivery of the IQI are maintained to ensure consistency in data storage, AMBI sensitivity categorisation and standardisation rules are reviewed periodically to ensure that assessments are applied consistently.

9.2.3 Continued collation of UK benthic data

To ensure the IQI is applied consistently across the UK, it is recommended that UK monitoring data continue to be collated. This is to ensure that aspects of the IQI that are dependent on data (for example, reference conditions, taxon lists) can be applied consistently and are relevant to all UK assessments (that is, all monitored habitats are sufficiently represented). It is also recommended that data collation is continued to enable the confidence in modelled parameters crucial for assessment (reference conditions and variability estimates) to be improved.

9.2.4 Review of $IQI_{v.IV}$ performance

The metrics within the IQI predominantly follow the principles of the Pearson–Rosenberg model of expected changes over an environmental disturbance gradient based on organic enrichment. Although the metrics incorporated within the IQI have been demonstrated to respond to other pressures such as hydromorphological changes (Desprez 2000, Newell et al. 2004, Muxika et al. 2005), analysis of the response of the IQI itself to additional identified pressures and across additional habitats is required to validate the method for wider applicability and to identify conditions where its use is appropriate. It is recommended that the performance of the IQI is assessed in terms of correlation to pressure data and variability under certain habitats and sampling methodology to identify the point at which the power to detect change in the IQI does not warrant the entailed sampling effort.

It is also recommended that seasonal effects on the IQI are investigated. The current restrictions on the Environment Agency WFD sampling window are based on the effect of seasonality on assemblage structure. Such seasonal differences in assemblage structure may not necessarily influence the metrics within the IQI (and therefore the IQI itself). Further understanding of the extent to which seasonality influences the IQI may enable the February to May sampling window to be extended.

Other metrics and indices that were not explored by Prior et al. (2004) in terms of their suitability for incorporation within the WFD classification tool may hold additional benefits to the IQI in terms of discriminating between natural and anthropogenic pressures (for example, the Benthic Quality Index, Quintino et al. 2006). Such metrics and indices may be useful as indicators to support WFD investigations.

9.2.5 Reference conditions

There are currently gaps in reference conditions for certain habitats and sample collection and processing methods. Data from these conditions need to be collated and incorporated into the approach to setting reference conditions to allow such data to be used for future classification.

It is also recommended that data should be added to the reference condition model developments as they become available. This will serve to improve the degree of confidence in the reference conditions and is of particular importance for habitats whereby current reference condition values are based on few data points (for example, habitats defined as gravely mud/sand).

9.2.6 Alternative approaches to model development

Regression analysis formed the basis of the development of the IQI and associated reference conditions. Improvements to the effectiveness of the classification tool and reference conditions may arise by exploring alternative approaches to model development such as distance-based linear models (DistLM), multiple discriminant analysis (MDA) and generalised additive mixed models (GAMM). It is recommended that any future developments explore the use of such approaches.

9.2.7 Variability

While the current report acknowledges that IQI variability may be influenced by a range of environmental conditions, the approach describes the differentiation of variability according to water body category (coastal or transitional) and aerial exposure (intertidal or subtidal), also acknowledging that differences may be attributed to different sample collection and processing. Continued analysis is recommended to further define variability estimates according to, for example, substratum so that sample effort (power to detect change) can be more accurately targeted based on different environmental conditions. There is also a need for further data analysis to establish additional components of variability relevant to the monitoring programmes (for example, between-year, within-station) of all UK competent authorities to ensure that variability estimates for the purposes of confidence of classification and power analysis are appropriate.

9.2.8 Confidence in reference conditions

It is recommended that the development/application of means for quantifying the variability in IQI reference conditions is incorporated within CofC and RoM estimates of assessments. In addition to improving the precision of CofC and RoM estimates, this would provide a basis for defining the environmental conditions and sample collection and processing methods where use of IQI may be considered to be both viable and logistically feasible.

References

AstraZeneca, 1980-1997. Ecological monitoring in the lower Tees estuary. Unpublished data.

AstraZeneca, 1999-2003. Marine monitoring for Cleveland Potash Ltd. Unpublished data.

Bald, J., Borja, Á., Muxika, I., Franco, J. and Valencia, V., 2005. Assessing reference conditions and physico-chemical status according to the European Water Framework Directive: a case-study from the Basque Country (Northern Spain). *Marine Pollution Bulletin,* 50 (12), 1508-1522.

Barbour, M.T., Stribling, J.B. and Karr, J.R., 1995. Multimetric approach for establishing biocriteria and measuring biological condition. In *Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making* (ed. W.S. Davis and T.P. Simon), pp. 63-77. Boca Raton, FL: Lewis Publishers.

Boaden, P.J.S and Seed, R., 1988. *An Introduction to Coastal Ecology.* Glasgow: Blackie and Son.

Borja, A. and Muxika, I., 2005. Guidelines for the use of AMBI (AZTI's marine biotic index) in the assessment of the benthic ecological quality. *Marine Pollution Bulletin*, 50 (7), 787-789.

Borja, A., Franco, J. and Pérez, V., 2000. A Marine Biotic Index to establish the ecological quality of soft-bottom benthos within European estuarine and coastal environments. *Marine Pollution Bulletin*, 40 (12), 1100-1114.

Borja, A., Franco, J. and Muxika, I., 2003. *Classification tools for marine ecological quality assessments: the usefulness of macrobenthic communities in an area affected by a submarine outfall.* ICES CM 2003/Session J-02, Tallinn, Estonia, 24-28 September 2003.

Borja, A., Josefson, A.B., Miles, A., Muxika, I., Olsgard, F., Phillips, G., Rodríguez, J.G. and Rygg, B., 2007. An approach to the intercalibration of benthic ecological status assessment in the North Atlantic ecoregion, according to the European Water Framework Directive. *Marine Pollution Bulletin,* 55 (1-6), 42-52.

Brown, A.R, and Shillabeer, N., 1997. Development of a biologically based environmental quality standard from a long-term benthic monitoring programme in the North Sea. *Oceanologica Acta*, 20(1), 275-282.

Carletti, A. and Heiskanen, A.-S. (eds.), 2009. *Water Framework Directive intercalibration technical report. Part 3: Coastal and transitional waters.* EUR 23838 En/3. Luxembourg: Office for Official Publications of the European Communities.

Cefas, 2009. *Green Book*. Monitoring Manual of the Clean Seas Environmental Monitoring Programme (CSEMP). Lowestoft: Centre for Environment, Fisheries and Aquaculture Science. Available from: http://www.cefas.defra.gov.uk/publications-and-data/scientific-series/green-book.aspx [Accessed 7 May 2013].

Clarke, K.R. and Warwick, R.M., 2001. *Change in Marine Communities: An Approach to Statistical Analysis and Interpretation*, 2nd ed. Plymouth: Primer-E.

COAST (WFD CIS Working Group 2.4), 2003. *Common Implementation Strategy for the Water Framework Directive Guidance Document No. 5. Transitional and coastal* *waters – typology, reference conditions and classification systems.* Luxembourg: Office for Official Publications of the European Communities.

Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement,* 20 (1), 37-46.

Connor, D.W., Brazier, D.P., Hill, T.O. and Northen, K.O., 1997a. *Marine Nature Conservation Review: Marine Biotope Classification for Britain and Ireland. Volume 1: Littoral Biotopes*. Peterborough: Joint Nature Conservation Committee.

Connor, D.W., Dalkin, M.J., Hill, T.O., Holt, R.H.F. and Sanderson, W.G., 1997b. *Marine Nature Conservation Review: Marine Biotope Classification for Britain and Ireland. Volume 2: Sublittoral Biotopes.* Peterborough: Joint Nature Conservation Committee.

Connor, D.W., Allen, J.H., Golding, N., Howell, K.L., Lieberknecht, L.M., Northen, K.O. and Reker, J.B., 2004. *The Marine Habitat Classification for Britain and Ireland*, version v04.05. Peterborough: Joint Nature Conservation Committee. Available from: http://www.jncc.gov.uk/MarineHabitatClassification [Accessed 7 May 2013].

Dauvin, J.C., 2007. Paradox of estuarine quality: benthic indicators and indices, consensus or debate for the future. *Marine Pollution Bulletin*, 55 (1-6), 271-281.

Davies, C.E. and Moss, D., 1998. *EUNIS Habitat Classification. Final Report to the European Topic Centre on Nature Conservation.* Copenhagen: European Environment Agency.

Davies, J., Baxter, J., Bradley, M., Connor, D., Khan, J., Murray, E., Sanderson, W., Turnbull, C. and Vincent, M. (eds.), 2001. *Marine Monitoring Handbook*. UK Marine SCAs Project. Available from: http://jncc.defra.gov.uk/page-2430 [Accessed 7 May 2013].

Davies, C.E., Moss, D. and Hill, M.O., 2004. *EUNIS Habitat Classification, Revised 2004.* Paris: European Topic Centre on Nature Protection and Biodiversity.

Desprez, M. 2000. Physical and biological impact of marine aggregate extraction along the French coast of the Eastern English Channel: short- and long-term post-dredging restoration. *ICES Journal of Marine Science,* 57 (5), 1428-1438.

Elliott, M. and Quintino, V., 2007. The Estuarine Quality Paradox, Environmental Homeostasis and the difficulty of detecting anthropogenic stress in naturally stressed areas. *Marine Pollution Bulletin*, 54 (6), 640-645.

Ellis, J. and Adriaenssens, V. 2006. *Uncertainty estimation for monitoring results by the WFD biological classification tools*. Bristol: Environment Agency

Environment Agency, 2012. *Water Framework Directive (WFD) sampling of macrobenthic invertebrates in transitional and coastal waters*. Operational Instruction 104_10. Bristol: Environment Agency.

Etter, R.J. and Grassle, J.F., 1992. Patterns of species diversity in the deep-sea as a function of sediment particle size. *Nature,* 360, 576-578.

European Commission, 2008. Commission Decision 2008/915/EC 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 Communities,* L332, 10.12.2008, 20-44.

European Parliament and Council of the European Union, 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. *Official Journal of the European Communities,* L327, 22.12.2000, 1-72.

Fisheries Research Services (Scotland), 1979-1988. Garroch Head sludge disposal ground survey. Unpublished data.

Fleiss, J.L. and Cohen, J. 1973. The equivalence of weighted Kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement,* 33 (3), 613-619.

Folk, R.L. 1954. The distinction between grain size and mineral composition in sedimentary rock nomenclature. *Journal of Geology,* 62 (4), 344-359.

Gibson, G.R., Bowman, M.L., Gerritsen, J. and Snyder, B.D., 2000. *Estuarine and coastal marine waters: bioassessment and biocriteria technical guidance.* EPA 822-B-00-024. Washington DC: US Environmental Protection Agency, Office of Water.

Grall, J., 2007. *A critical evaluation of the benthic indices used to assess ecological status of coastal waters within the Water Framework Directive*. Presentation to EC-funded workshop on 'Pollutant monitoring and ecological impact assessment following accidental oil and other chemical spills in marine waters', Cedre, Brest, France, 9-11 October 2007.

Grall, J. and Glémarec, M., 1997. Using biotic indices to estimate macrobenthic community perturbations in the Bay of Brest. *Estuarine, Coastal and Shelf Science,* 44 (Suppl. 1), 43-53.

Gray, J.S., Poore, G C.B., Ugland, K.I., Wilson, R.S., Olsgard, F. and Johannessen, O., 1997. Coastal and deep-sea benthic diversities compared. *Marine Ecology Progress Series,* 159, 97-103.

Hayward, P.J. and Ryland, J.S., 1995. *Handbook of the Marine Fauna of the British Isles and Northwest Europe.* Oxford: Oxford University Press.

Hily, C., 1984. *Variabilite de la macrofaune benthique dans les milieux hypertrophiques de la Rade de Brest.* PhD thesis, University of Bretagne Occidentale.

Hily, C., Le Bris, H. and Glémarec, M., 1986. Impacts biologiques des emissaires urbains sur les ecosystemes benthiques. *Oceanis*, 12 (6), 419-426.

ISO (International Organization for Standardization), 2005. ISO 2005 16665:2005 *Water quality – Guidelines for quantitative sampling and sample processing of marine soft-bottom macrofauna* [under review]. Geneva: ISO.

Landis, J.R. and Koch, G.G., 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33 (1), 159-174.

Lewis, F.G. and Stonor, A.W., 1981. An examination of methods for sampling macrobenthos in seagrass meadows. *Bulletin of Marine Science*, 31 (1), 116-124.

Majeed, S.A., 1987. Organic matter and biotic indices on the beaches of North Brittany. *Marine Pollution Bulletin*, 18 (9), 490-495.

Millward, R.N. and Grant, A., 1995. Assessing the impact of copper on nematode communities from a chronically metal-enriched estuary using pollution-induced community tolerance. *Marine Pollution Bulletin*, 30 (11), 701-706.

Monserud, R. and Leemans, R., 1992. Comparing global vegetation maps with the Kappa statistic. *Ecological Modelling*, 62 (4), 275-293.

Muxika, I., Borja, Á. and Bonne, W., 2005. The suitability of the marine biotic index (AMBI) to new impact sources along European coasts. *Ecological Indicators*, 5 (1), 19-31.

Newell, R.C., Seiderer, L.J., Simpson, N.M. and Robinson, J.E., 2004. Impacts of marine aggregate dredging on benthic macrofauna off the south coast of the United Kingdom. *Journal of Coastal Research*, 20 (1), 115-125.

Pearson, T.H. and Rosenberg, R., 1978. Macrobenthic succession in relation to organic enrichment and pollution of the marine environment. *Oceanography and Marine Biology Annual Review*, 16, 229-311.

Prior, A., Miles, A.C., Sparrow, A.J. and Price, N., 2004. *Development of a classification scheme for the marine benthic invertebrate component, Water Framework Directive. Phases I & II – Transitional and coastal waters.* R&D Interim Technical Report E1-116, E1-132. Bristol: Environment Agency.

Proudfoot, R.K., Elliott, M., Dyer, M.F., Barnett, B.E., Allen, J.H., Proctor, N.L., Cutts, N.D., Nikitik, C., Turner, G., Breen, J., Hemmingway, K.L. and Mackie, T., 1997. Collection and processing of macrobenthic samples from soft sediments; a best practice review. Proceedings of the Humber Benthic Field Methods Workshop, University of Hull, 1997.

Quintino, V., Elliott, M. and Rodrigues, A.M., 2006. The derivation, performance and role of univariate and multivariate indicators of benthic change: case studies at differing spatial scales. *Journal of Experimental Marine Biology and Ecology*, 330 (1), 368-382.

REFCOND (WFD CIS Working Group 2.31 Reference Conditions for Inland Surface Waters), 2003. *Guidance on establishing reference conditions and ecological status class boundaries for inland surface waters.* Brussels: European Commission.

Rygg, B., 2002. *Indicator species index for assessing benthic ecological quality in marine waters of Norway*. Report SNO 4548-2002. Oslo: Norwegian Institute for Water Research.

Simboura, N. and Zenetos, A., 2002. Benthic indicators to use in ecological quality classification of Mediterranean soft bottom marine ecosystems, including a new biotic index. *Mediterranean Marine Science*, 3 (2), 77-111.

Solheim, A., Andersen, T., Brettum, P., Bækken, T., Bongard, T., Moy, F., Kroglund, T., Olsgard, F., Rygg, B. and Oug, E., 2004. *BIOKLASS – Classification of ecological status in Norwegian water bodies: relevant criteria and possible boundary values between good and moderate ecological status for selected elements and pressures.* Report No 4860. Oslo: Norwegian Institute for Water Research.

Symposium on the Classification of Brackish Waters, 1959. Final resolution. The Venice system for the classification of brackish waters according to salinity. *Archivio di Oceanografia e Limnologia*, 11 (Suppl.), 243-248.

UKTAG (United Kingdom Technical Advisory Group on the Water Framework Directive), 2003a. *Guidance on typology for coastal and transitional waters of the UK and Republic of Ireland*. Edinburgh: SNIFFER. Available from: http://www.wfduk.org/tagged/typology [Accessed 7 May 2013].

UKTAG (United Kingdom Technical Advisory Group on the Water Framework Directive), 2003b. *Guidance on general principles for pressures & impacts analysis.* Edinburgh: SNIFFER. Available from: http://www.wfduk.org/resources%20/guidance-general-principles-pressures-impacts-analysis-0 [Accessed 7 May 2013].

UKTAG (United Kingdom Technical Advisory Group on the Water Framework Directive), 2008. *Coastal water assessment method benthic invertebrate fauna: invertebrates in soft sediments (Infaunal Quality Index).* Edinburgh: SNIFFER.

Vincent, C., Heinrich, H., Edwards, A., Nygaard, K. and Haythornwaite, J., 2002. *Guidance on typology, reference conditions and classification system for transitional and coastal waters.* Final draft. Produced by CIS Working Group 2.4 (COAST), Common Implementation Strategy of the Water Framework Directive. Brussels: European Commission.

Warwick, R.M., Ashman, C.M., Brown, A.R., Clarke, K.R., Dowell, B., Hart, B., Lewis, R.E., Shillabeer, N., Somerfield, P.J. and Tapp, J.F., 2002. Inter-annual changes in the biodiversity and community structure of the macrobenthos in Tees Bay and the Tees estuary, UK, associated with local and regional environmental events. *Marine Ecology Progress Series*, 234, 1-13.

Washington, H.G., 1984. Diversity, biotic and similarity Indices: a review with special relevance to aquatic ecosystems. *Water Research,* 18 (6), 653-694.

Weisberg, S.B., J. Ananda Ranasinghe, Dauer, D.M., Schaffner, L.C., Diaz, R.J. and Frithsen, J.B. 1997. An Estuarine Benthic Index of Biotic Integrity (B-IBI) for Chesapeake Bay Estuaries. *Estuaries*, 20, 149-158.

Wright, J.F., Sutcliffe, D.W. and Furse, M.T. (eds.), 2000. *Assessing the Biological Quality of Fresh Waters: RIVPACS and Other Techniques.* Ambleside, Cumbria: The Freshwater Biological Association.

Zar, J.H., 1984. *Biostatistical Analysis*. 2nd Edition. Englewood Cliffs, NJ: Prentice-Hall.

List of abbreviations

AMBI	AZTI Marine Biological Index
ANOVA	analysis of variance
AQC	analytical quality control
AZTI	Marine and Food Technological Centre
CCW	Countryside Wales
Cefas	Centre for the Environment, Fisheries and Aquaculture Science
CEN	European Committee for Standardisation (Comité Européen de Normalisation)
CI	confidence interval
CIS	Common Implementation Strategy
COAST	Common Implementation Strategy Transitional and Coastal Working Group
CofC	confidence of class
CSEMP	Clean Seas Environmental Monitoring Programme (formerly NMMP)
CW	coastal water
Defra	Department for Environment, Food and Rural Affairs
DKI	Danish Quality Index
ECOSTAT	Ecological Status Working Group
EG	Ecological Group
EQR	Ecological Quality Ratio
EUNIS	European Nature Information System
FRS	Fisheries Research Services (Scotland)
GEP	Good Ecological Potential
GES	Good Ecological Status
HMWB	heavily modified water body
ICES	International Council for the Exploration of the Seas
IECS	Institute of Estuarine and Coastal Studies
IQI	Infaunal Quality Index
ISI	Indicator Species Index
ISO	International Organization for Standardization
ITI	Infaunal Trophic Index
JAMP	Joint Assessment Monitoring Programme

JNCC	Joint Nature Conservancy Council
M-AMBI	Multivariate AZTI Marine Biotic Index
MarLIN	The Marine Life Information Network for Britain and Ireland
MBITT	Marine Benthic Invertebrate Task Team
MI RoI	Marine Institute of the Republic of Ireland
MNCR	Marine Nature Conservation Review
MTT	Marine Task Team
NEA	North East Atlantic
NEAGIG	North East Atlantic Geographical Intercalibration Group
NKI	Norwegian Quality Index
NIEA	Northern Ireland Environment Agency
NMBAQC	National Marine Biological Analytical Quality Control
NMMP	National Marine Monitoring Programme
OSPAR	Oslo and Paris Conventions
P-BAT	Portuguese Benthic Assessment Tool
PC1	first principal component
PCA	principal components analysis
PRIMER©	Plymouth Routines in Multivariate Ecological Research
PSA	particle size analysis
QA	quality assurance
R&D	research and development
RBMP	River Basin Management Plan
RoI	Republic of Ireland
RoM	risk of miscalculation
SD	standard deviation
SE	standard error
SEPA	Scottish Environment Protection Agency
SNIFFER	Scottish and Northern Ireland Forum for Environmental Research
SOP	Standard Operating Procedure
TraC	transitional and coastal
TW	transitional water
UKTAG	United Kingdom Technical Advisory Group
WFD	Water Framework Directive
WRc	Water Research Centre

Appendix A Family level truncation codes

Table A1 Family level truncation codes following revision of the IQI truncation rules in 2008 (0 = exclude from analysis, 1 = enumerate as one individual per sample, 2 = retain unaltered)

Family	Truncation code	Family	Truncation code
Acanthochitonidae	2	Aoridae	2
Acanthonotozomatidae	2	Aphelocheridae	2
Aclididae	2	Aphididae	0
Acoetidae	2	Aphroditidae	2
Acrocirridae	2	Apistobranchidae	2
Acteonidae	2	Aplysiidae	2
Actiniidae	2	Aporrhaidae	2
Adeonidae	1	Apseudidae	2
Adeorbidae	2	Arachnidiidae	1
Aegidae	2	Arcidae	2
Aeolidiidae	2	Arcticidae	2
Aeteidae	1	Arcturidae	2
Agelenidae	2	Arenicolidae	2
Aglaopheniidae	1	Argissidae	2
Akeridae	2	Arminidae	2
Alcyonidiidae	1	Artotrogidae	0
Alcyoniidae	1	Ascidicolidae	0
Alderiidae	2	Ascidiidae	2
Alpheidae	2	Asellidae	2
Ammotheidae	2	Aspidosiphonidae	2
Ampeliscidae	2	Astartidae	2
Ampharetidae	2	Asteriidae	2
Amphilepidae	2	Asterinidae	2
Amphilochidae	2	Asthenognathidae	2
Amphinomidae	2	Astropectinidae	2
Amphiporidae	2	Atelecyclidae	2
Amphiuridae	2	Atylidae	2
Ampithoidae	2	Axiidae	2
Anarthruridae	2	Axinellidae	1
Ancylidae	2	Baetidae	2
Annectocymidae	1	Balanidae	1
Anomiidae	2	Barentsiidae	1
Antedonidae	2	Barleeidae	2
Anthuridae	2	Bathyporeiidae	2

Family	Truncation code
Beaniidae	1
Bodotriidae	2
Bonnelliidae	2
Bopyridae	0
Bougainvilliidae	1
Branchiostomidae	2
Brissidae	2
Buccinidae	2
Bugulidae	1
Bythocytheridae	2
Cadlinidae	2
Caecidae	2
Caenidae	2
Calanidae	0
Callianassidae	2
Calliopiidae	2
Callipallenidae	2
Calloporidae	1
Calocarididae	2
Calyptraeidae	2
Campanulariidae	1
Campanulinidae	1
Cancridae	2
Candidae	1
Capitellidae	2
Caprellidae	2
Capulidae	2
Cardiidae	2
Carinomidae	2
Caryophyllaeidae	2
Caryophylliidae	2
Cavernulariidae	2
Cellariidae	1

Family	Truncation code
Celleporidae	1
Cephalothricidae	2
Ceratopogonidae	0
Cerebratulidae	2
Cerianthidae	2
Cerithiidae	2
Cerithiopsidae	2
Chaetodermatidae	2
Chaetopteridae	2
Chalinidae	1
Cheluridae	2
Chironomidae	2
Chorizoporidae	1
Chrysopetalidae	2
Chthamalidae	1
Cionidae	2
Cirolanidae	2
Cirratulidae	2
Clavidae	1
Clionidae	1
Cochliopidae	2
Coenagriidae	2
Colomastigidae	2
Corbiculidae	2
Corbulidae	2
Corellidae	2
Corixidae	0
Corophiidae	2
Corymorphidae	2

Family	Truncation code	Family	Truncation code
Corynidae	1	Diastylidae	2
Coryphellidae	2	Didemnidae	1
Corystidae	2	Diogenidae	2
Cossuridae	2	Diosaccidae	0
Crangonidae	2	Dolichopodidae	0
Crangonyctidae	2	Donacidae	2
Craniidae	2	Dondersiidae	2
Cressidae	2	Dorididae	2
Cribrilinidae	1	Dorvilleidae	2
Crisiidae	1	Dotidae	2
Cryptosulidae	1	Dreissenidae	2
Ctenodiscidae	2	Dromiidae	2
Ctenodrilidae	2	Dugesiidae	2
Cucumariidae	2	Dysideidae	1
Culicidae	0	Dytiscidae	2
Curculionidae	2	Ebalidae	2
Cuspidariidae	2	Echinidae	2
Cylichnidae	2	Echiuridae	2
Cylindroleberidae	2	Edwardsiidae	2
Cyprididae	2	Electridae	1
Cytherideidae	2	Eleutherocarpidae	2
Cytheruridae	2	Ellobiidae	2
Dajidae	0	Elmidae	2
Dendrocoelidae	2	Enchelidiidae	2
Dendronotidae	2	Enchytraeidae	2
Dentaliidae	2	Endeidae	2
Dexaminidae	2	Enoplidae	2
Diaphanidae	2	Entalinidae	2
Diastoporidae	1	Ephemeridae	2

Family	Truncation code	Family	Truncation code
Epimeriidae	2	Glossidae	2
Epitoniidae	2	Glossiphoniidae	2
Epizoanthidae	1	Glyceridae	2
Erpobdellidae	2	Glycymerididae	2
Escharellidae	1	Gnathiidae	2
Eubranchidae	2	Golfingiidae	2
Eucheilotidae	1	Gonactiniidae	2
Eucrateidae	1	Goneplacidae	2
Eudendriidae	1	Goniadidae	2
Eulimidae	2	Goniodorididae	2
Eunicidae	2	Grantiidae	1
Euphausiidae	0	Grapsidae	2
Euphrosinidae	2	Halacaridae	2
Eurycopidae	2	Halcampidae	2
Euryleptidae	2	Haleciidae	1
Eusiridae	2	Halichondriidae	1
Exochellidae	1	Haliplidae	2
Facelinidae	2	Haloclavidae	2
Fecampiidae	2	Haminoeidae	2
Fibulariidae	2	Hanleyidae	2
Fissurellidae	2	Harrimanidae	2
Flabelligeridae	2	Haustoriidae	2
Flustrellidridae	1	Hemiasterellidae	1
Flustridae	1	Hesionidae	2
Formicidae	0	Hiatellidae	2
Galatheidae	2	Hippolytidae	2
Galeommatidae	2	Hippoporidridae	1
Gammarellidae	2	Hippoporinidae	1
Gammaridae	2	Hippothoidae	1

Family	Truncation code	Family	Truncation code
Holothuriidae	2	Leptochitonidae	2
Hormathiidae	2	Leptocytheridae	2
Hyalidae	2	Leptognathiidae	2
Hydractiniidae	1	Leptonidae	2
Hydridae	1	Leptoplanidae	2
Hydrobiidae	2	Leuconiidae	2
Hydroptilidae	2	Leucosiidae	2
Hyperiidae	0	Leucosolenidae	1
Idoteidae	2	Leucothoidae	2
Ilyarachnidae	2	Lichenoporidae	1
Incertae sedis	2	Ligiidae	2
Iravadiidae	2	Liljeborgiidae	2
Isaeidae	2	Limapontiidae	2
Ischnochitonidae	2	Limidae	2
Ischyroceridae	2	Limifossoridae	2
Janiridae	2	Limnoriidae	2
Kelliellidae	2	Lineidae	2
Kelliidae	2	Littorinidae	2
Lacunidae	2	Loliginidae	0
Lacydoniidae	2	Lomanotidae	2
Lafoeidae	1	Longipediidae	0
Lamellariidae	2	Lottiidae	2
Lampropidae	2	Lovenellidae	1
Laomediidae	2	Loveniidae	2
Laophontidae	0	Loxosomatidae	2
Lasaeidae	2	Lucinidae	2
Lepetidae	2	Luidiidae	2
Lepidomeniidae	2	Lumbricidae	2
Leptoceridae	2	Lumbriculidae	2
		Lumbrineridae	2
		Lymnaeidae	2
		Lyonsiidae	2
		Lysianassidae	2
		Mactridae	2

Family	Truncation code	Family	Truncation code
Magelonidae	2	Neomeniidae	2
Majidae	2	Nephropidae	2
Malacobdellidae	2	Nephtyidae	2
Maldanidae	2	Nereididae	2
Malletiidae	2	Neritidae	2
Megaluropodidae	2	Nolellidae	1
Melinnacheridae	0	Notodelphyidae	0
Melitidae	2	Nototanaidae	2
Melphidippidae	2	Nuculanidae	2
Membraniporidae	1	Nuculidae	2
Metridiidae	2	Nymphonidae	2
Microcionidae	1	Oedicerotidae	2
Microporellidae	1	Oenonidae	2
Mimosellidae	1	Omalogyridae	2
Molgulidae	2	Onchidorididae	2
Montacutidae	2	Oncholaimidae	2
Munnidae	2	Oncousoeciidae	1
Munnopsidae	2	Onuphidae	2
Muricidae	2	Opheliidae	2
Mycalidae	1	Ophiactidae	2
Myidae	2	Ophiocomidae	2
Myriotrochidae	2	Ophiotrichidae	2
Mysidae	0	Ophiuridae	2
Mytilicolidae	0	Oplophoridae	0
Mytilidae	2	Orbiniidae	2
Myxillidae	1	Ostreidae	2
Myxinidae	0	Oweniidae	2
Naididae	2	Paguridae	2
Nannastacidae	2	Palaemonidae	2
Nassariidae	2	Pandalidae	2
Naticidae	2	Pandeiidae	1
Nebaliidae	2	Pandoridae	2
Nemouridae	2	Paradoxostomatidae	2
Neoleptonidae	2	Paramunnidae	2

Family	Truncation code	Family	Truncation code
Paraonidae	2	Pilargidae	2
Paratanaidae	2	Pilumnidae	2
Parazoanthidae	1	Pinnotheridae	2
Pardaliscidae	2	Pirimelidae	2
Parechinidae	2	Piscicolidae	0
Pariambidae	2	Pisidiidae	2
Patellidae	2	Pisionidae	2
Pectinariidae	2	Planariidae	2
Pectinidae	2	Planorbidae	2
Pedicellinidae	1	Pleurobrachiidae	0
Peltidiidae	0	Pleurobranchidae	2
Penetrantiidae	1	Pleurogonidae	2
Pennatulidae	2	Pleustidae	2
Periplomatidae	2	Plumulariidae	1
Perophoridae	1	Podoceridae	2
Petricolidae	2	Podonidae	0
Pharidae	2	Poecilochaetidae	2
Phascoliidae	2	Polyceridae	2
Phascolosomatidae	2	Polyclinidae	1
Phasianellidae	2	Polygordiidae	2
Phialellidae	1	Polymastiidae	1
Philinidae	2	Polynoidae	2
Philinoglossidae	2	Polyposthiidae	2
Philomedidae	2	Pontoporeiidae	2
Phliantidae	2	Porcellanidae	2
Pholadidae	2	Poromyidae	2
Pholoidae	2	Portunidae	2
Phoronidae	2	Priapulidae	2
Phoxichilidiidae	2	Processidae	2
Phoxocephalidae	2	Propilidiidae	2
Phryxidae	2	Protodrilidae	2
Phtisicidae	2	Protodriloidae	2
Phyllodocidae	2	Psammobiidae	2
Physidae	2	Psammodrilidae	2

Family	Truncation code	Family	Truncation code
Pseudocumatidae	2	Sergestidae	0
Psolidae	2	Serpulidae	2
Psychodidae	2	Sertulariidae	1
Psychomyiidae	2	Sigalionidae	2
Ptychoderidae	2	Sipunculidae	2
Pulsellidae	2	Skeneidae	2
Pycnogonidae	2	Skeneopsidae	2
Pyramidellidae	2	Smittinidae	1
Pyuridae	2	Solasteridae	2
Raspailiidae	1	Solenidae	2
Retusidae	2	Spadellidae	2
Rhopalomeniidae	2	Spatangidae	2
Ringiculidae	2	Sphaeridae	2
Rissoellidae	2	Sphaerodoridae	2
Rissoidae	2	Sphaeromatidae	2
Runcinidae	2	Spionidae	2
Sabellariidae	2	Spirorbidae	2
Sabellidae	2	Staphylinidae	2
Sabelliphilidae	0	Stegocephalidae	2
Saccocirridae	2	Stenothoidae	2
Sacculinidae	0	Sternaspidae	2
Sagartiidae	2	Stiligeridae	2
Sagittidae	0	Strongylocentrotidae	2
Sareptidae	2	Styelidae	1
Sarsiellidae	2	Suberitidae	1
Scalibregmatidae	2	Sycettidae	1
Scalpellidae	1	Syllidae	2
Scaphandridae	2	Synaptidae	2
Schizasteridae	2	Synopiidae	2
Schizoporellidae	1	Taeniopterygidae	2
Scrobiculariidae	2	Talitridae	2
Scrupariidae	1	Tanaidae	2
Semelidae	2	Tellinidae	2
Sepiolidae	2	Terebellidae	2

Family	Truncation code	Family	Truncation code
Tergipedidae	2	Virgulariidae	2
Tetrastemmatidae	2	Walkeriidae	1
Teuchoporidae	1	Xanthidae	2
Thiidae	2	Ypsilothuriidae	2
Thraciidae	2		
Thyasiridae	2		
Tipulidae	2		
Tornidae	2		
Trachyleberididae	2		
Trichobranchidae	2		
Triphoridae	2		
Triticellidae	1		
Tritoniidae	2		
Triviidae	2		
Trochidae	2		
Trochochaetidae	2		
Tubificidae	2		
Tubulanidae	2		
Tubulariidae	1		
Tubuliporidae	1		
Turridae	2		
Turritellidae	2		
Turtoniidae	2		
Typhlotanaidae	2		
Ulmaridae	0		
Umbonulidae	1		
Ungulinidae	2		
Upogebiidae	2		
Urothoidae	2		
Valvatidae	2		
Veneridae	2		
Verrucidae	1		
Vesiculariidae	1		
Victorellidae	1		

Appendix B Expanded normative definitions

Table B1 Coastal waters (EUNIS Habitat A.4 sublittoral sediments)

Quality status	Normative definition	Expanded interpretation

High	Level of diversity and abundance of invertebrate taxa is within the range normally associated with undisturbed conditions.
	
	All disturbance-sensitive taxa associated with undisturbed conditions are present.
		Invertebrate community shows no anthropogenic impact.
		• Species richness and diversity high (for example, number of species, Shannon, Fisher, Margalef and Brillouin diversity indices)
		• Evenness high (Heip and Pielou indices); abundance ratio (abundance/number of taxa) low
		• Taxonomic range high (taxonomic diversity, distinctness, and breadth indices)
		• Community abundances (assessed by AMBI) – normal, unpolluted:
		  • sensitive taxa (EGI) of dominant abundance
		  • indifferent and tolerant taxa (EGII and EGIII) absent or of sub-dominant abundance
		  • opportunistic taxa (EGIV) absent or of negligible abundance
		  • indicator taxa (EGV) absent or of negligible abundance
		• Trophic structure (assessed by ITI) – normal:
		  • dominated by water column and interface detritus feeders
		• Abundance of important characterising, structural or functional species unimpacted (for example, seapens or burrowing decapods, large bivalves)

Good	Level of diversity and abundance of invertebrate taxa is slightly outside the range associated with the type-specific	**Invertebrate community shows slight anthropogenic impact.** • Species richness and diversity slightly reduced (for example, Shannon, Fisher,

Quality status	Normative definition	Expanded interpretation

	conditions.
	
	Most of the sensitive taxa of the type-specific conditions are present.
		Margalef and Brillouin diversity indices)
		
		• Evenness slightly reduced (Heip and Pielou indices); abundance ratio slightly elevated
		
		• Taxonomic range slightly reduced (taxonomic diversity, distinctness, and breadth indices)
		
		• Community abundances (assessed by AMBI) – slightly unbalanced, slightly polluted:
		
		  • sensitive taxa (EGI) abundance may range from high sub-dominant to absent
		
		  • indifferent taxa (EGII) of low sub-dominant abundance
		
		  • tolerant taxa (EGIII) of dominant abundance
		
		  • opportunistic taxa (EGIV) and indicator taxa (EGV) abundance may range from negligible or low to equi-abundance with indifferent taxa
		
		• Trophic structure (assessed by ITI) – normal or slightly changed:
		
		  • dominated by detritus and deposit feeders
		
		• Abundance of important characterising, structural, or functional species slightly reduced (for example, seapens or burrowing decapods, large bivalves)

Moderate	Level of diversity and abundance of invertebrate taxa is moderately outside the range associated with the type-specific conditions. Taxa indicative of pollution are present. Many of the sensitive taxa of the type-specific communities are absent.	**Invertebrate community shows moderate anthropogenic impact.** • Species richness and diversity moderately reduced (for example, number of species, Shannon, Fisher, Margalef and Brillouin diversity indices) • Evenness moderately reduced (Heip and Pielou indices); abundance ratio moderately elevated • Taxonomic range moderate reduced. (taxonomic diversity, distinctness, and breadth indices) • Community abundances (assessed by AMBI) – transitional unbalanced to

Quality status	Normative definition	Expanded interpretation
		moderately polluted: • sensitive taxa (EGI) of negligible abundance or absent • indifferent taxa (EGII) of low sub-dominant abundance • tolerant taxa (EGIII), opportunistic taxa (EGIV) and indicator taxa (EGV) co- dominate the abundance • Trophic structure (assessed by ITI) – shows moderate change: • dominated by interface deposit feeders • Abundance of important characterising, structural or functional species moderately reduced. Some key species of negligible abundance or absent (for example, seapens or burrowing decapods, large bivalves)
Poor	Waters showing evidence of major alterations to the values of the biological quality elements for the surface water body type and in which the relevant biological communities deviate substantially from those normally associated with the surface water body type under undisturbed conditions.	**Invertebrate community shows major anthropogenic impact.** • Species richness and diversity shows major reduction (for example, number of species, Shannon, Fisher, Margalef and Brillouin diversity indices) • Evenness shows major reduction (Heip and Pielou indices); abundance ratio shows major elevation • Taxonomic range shows major reduction. (taxonomic diversity, distinctness, and breadth indices) • Community abundances (assessed by AMBI) – transitional moderately to heavily polluted: • sensitive and indifferent taxa (EGI and EGII) of negligible abundance or absent • tolerant taxa (EGIII) of sub-dominant abundance • opportunistic taxa (EGIV) and indicator taxa (EGV) co-dominate the abundance. • Trophic structure (assessed by ITI) –

Quality status	Normative definition	Expanded interpretation
		shows major change or degradation: • dominated by interface and subsurface deposit feeders • Abundance of important characterising, structural, or functional species shows major reduction. Many key species of negligible abundance or absent (for example, seapens or burrowing decapods, large bivalves)
Bad	Waters showing evidence of severe alterations to the values of the biological quality elements for the surface water body type and in which large portions of the relevant biological communities normally associated with the surface water body type under undisturbed conditions are absent .	**Invertebrate community shows severe anthropogenic impact.** • Species richness and diversity shows severe reduction (for example, number of species, Shannon, Fisher, Margalef and Brillouin diversity indices) • Evenness shows severe reduction (Heip and Pielou indices); abundance ratio shows severe elevation • Taxonomic range severely reduced (Taxonomic diversity, distinctness, and breadth indices) • Community abundances (assessed by AMBI) – very heavily or extremely polluted: • azoic or if fauna present: – sensitive, indifferent and tolerant Taxa (EGI, EGII and EGIII) absent – opportunistic taxa (EGIV) of sub-dominant abundance – indicator taxa (EGV) of dominant abundance • Trophic structure (assessed by ITI) – shows severe degradation: • dominated by subsurface deposit feeders, or azoic • All important characterising, structural, or functional species of negligible abundance or absent (for example, seapens or burrowing decapods, large bivalves)

I

We are The Environment Agency. It's our job to look after your environment and make it **a better place** – for you, and for future generations.

Your environment is the air you breathe, the water you drink and the ground you walk on. Working with business, Government and society as a whole, we are making your environment cleaner and healthier.

The Environment Agency. Out there, making your environment a better place.

Would you like to find out more about us, or about your environment?

Then call us on

03708 506 506 (Mon-Fri 8-6)

Calls to 03 numbers cost the same as calls to standard geographic numbers (i.e. numbers beginning with 01 or 02).

email

enquiries@environment-agency.gov.uk

or visit our website

www.environment-agency.gov.uk

incident hotline 0800 80 70 60 (24hrs)
floodline 0845 988 1188

Environment first: Are you viewing this on screen? Please consider the environment and only print if absolutely necessary. If you are reading a paper copy, please don't forget to reuse and recycle if possible.