ENVIRONMENT AGENCY

CONFIDENCE OF CLASS FOR WFD MARINE PLANT TOOLS

WRc Ref: EA7954
22/11/2013

CONFIDENCE OF CLASS FOR WFD MARINE PLANT TOOLS

Report No.:	EA7954
Date:	12/03/2009
Authors:	Andrew Davey
Contract Manager:	Ian Codling
Contract No.:	15144-0

Updated by: Keith McGruer, Environment Agency

Date: 22/11/2013

Any enquiries relating to this report should be referred to the authors at the following address:

WRc Swindon, Frankland Road, Blagrove, Swindon, Wiltshire, SN5 8YF.
Telephone: + 44 (0) 1793 865000 Fax: + 44 (0) 1793 865001
Website: www.wrcplc.co.uk

This document has been produced by WRc plc.

CONTENTS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The Water Framework Directive requires the Environment Agency to classify all surface waterbodies into one of five status classes: High, Good, Moderate, Poor or Bad. In addition, the Agency is required to report the level of confidence associated with waterbody classifications.

In an ideal world of comprehensive monitoring data containing no errors, waterbodies would always be assigned to their true class with 100% confidence, but in reality estimates of the truth based on monitoring are always subject to error. Understanding and managing the risk of misclassification as a result of uncertainties in the results of monitoring is important on two counts; first, because of the potential to fail to act in cases where a waterbody has been wrongly classified as being of better status than it is, and secondly because of the risk of wasting resources on waterbodies that have been wrongly classified as worse than they are.

WRc was commissioned by the Agency to develop a suite of tools to calculate the confidence of classification associated with assessment of the status of marine plant communities.

Three spreadsheet tools were developed:

1. CAPTAIN – to calculate CofC for the opportunistic macroalgae assessment tool; and
2. PUGWASH – to calculate CofC for the phytoplankton assessment tool;
3. PIRATES – to calculate CofC for the rocky shore reduced species list (RSL) assessment tool.

This report summarises the work undertaken and documents the statistical methodology used in these tools.

1. INTRODUCTION

The Water Framework Directive (WFD) requires the Environment Agency (hereafter 'the Agency') to classify all surface waterbodies into one of five status classes: High, Good, Moderate, Poor or Bad. In addition, the Agency is required to report the level of confidence associated with waterbody classifications.

In an ideal world of comprehensive monitoring data containing no errors, waterbodies would always be assigned to their true class with 100% confidence. But estimates of the truth based on monitoring are subject to error because monitoring is not done everywhere and all the time, and because monitoring systems, equipment and people are less than perfect. Understanding and managing the risk of misclassification as a result of uncertainties in the results of monitoring is important on two counts; first, because of the potential to fail to act in cases where a waterbody has been wrongly classified as being of better status than it is, and secondly because of the risk of wasting resources on waterbodies that have been wrongly classified as worse than they are.

The ecological status of each waterbody is assessed using one or more biological quality elements, each of which yields an Environmental Quality Ratio (EQR) score between 0 (Bad status) and 1 (High status). The EQR for each biological quality element in each waterbody must be accompanied by a measure of the confidence of class (CofC), and also the confidence that the EQR class is of Good or High status (CofGorH). A preliminary methodology for calculating CofC for the three marine plant tools was developed by WRc in 2007, and in October 2008 WRc was commissioned by the Agency to refine and complete the CofC calculation methodology.

Three spreadsheet tools were developed during this second phase of work:

1. CAPTAIN – to calculate CofC for the opportunistic macroalgae assessment tool; and

2. PUGWASH – to calculate CofC for the phytoplankton assessment tool;

3. PIRATES – to calculate CofC for the rocky shore reduced species list (RSL) assessment tool.

This report briefly summarises the work undertaken and documents the statistical methodology used in the three spreadsheet tools.

The remainder of the report is divided in four sections: Section 2 sets out the background to this project and summarises the general approach taken towards assessing uncertainty in these multi-metric tools, while Sections 3, 4 and 5 detail the statistical methodology used in CAPTAIN, PUGWASH and PIRATES, respectively.

2. BACKGROUND AND APPROACH

2.1 Background

The ecological status of transitional (estuarine) and coastal waterbodies is assessed using a variety of biological quality elements, including macroalgae and phytoplankton.

For each quality element, the ecological status of the waterbody is measured by an Ecological Quality Ratio (EQR), which comprises one or more sub-metrics that measure different aspects of biological community. For example, the phytoplankton assessment tool has three sub-metrics: chlorophyll-a concentration, elevated counts and seasonal succession. EQRs take a value between 0 and 1, and this range is split into five status classes (Bad, Poor, Moderate, Good and High).

For classification purposes, the estimated EQR is translated directly into a face value class. However, because it is not possible to survey biological community across whole waterbody continuously throughout whole reporting period, there will always be some sampling error, which will lead to uncertainty in the estimate of the EQR. This uncertainty can be quantified as the expected difference between the observed EQR and the true underlying EQR, which can then be used to calculate the probability of the waterbody being in each of the five status classes. From this it is possible to determine the most probable class (the one with the highest probability) and state what level of confidence we have that the true status is good or better, and moderate or worse.

2.2 Previous work

This project builds upon and integrates a number of spreadsheet tools developed by WRc for the Agency under previous contracts.

Ellis & Adriaenssens (2006) developed a spreadsheet tool called SDvMean to derive estimates of the combined spatial and temporal variability in EQR results in waterbodies that lacked replicate surveys. Specifically, it used EQR data from other waterbodies to fit a power curve relating the mean EQR to the standard deviation of the individual EQR results. The power curve is then used to predict the likely variability in waterbodies that have just a single EQR result with which to estimate the mean. Further details are given in Appendix A.

Ellis & Adriaenssens (2006) also developed a generic approach to convert uncertainty in the estimated EQR into a confidence of class. It uses a logit transformation to ensure that the calculations are constrained to lie within the 0 to 1 EQR range, and then calculates the probability that the true EQR lies in each of the five status classes. This approach was first implemented in the CofC.xls spreadsheet tool. Further details are given in Appendix B.

2.3 Stance towards sub-metric uncertainty

All three biological tools considered in this study calculate an EQR that comprises multiple sub-metrics. Two contrasting approaches were identified to deal with uncertainty in cases where multiple sub-metrics are combined, typically by averaging, to give an overall EQR: a ‘*short-cut*’ and a ‘*bottom-up*’ approach.

The short-cut approach focuses on the variation *among* the sub-metric EQRs when computing the CofC. It assumes the sub-metrics measure the same aspect of the biological community and that, given perfect information, the sub-metrics would converge at the same EQR value. This has the useful consequence that one need not compute the uncertainty within each sub-metric, because it is intrinsically part of the variation among the sub-metric EQRs.

By contrast, the bottom-up approach focuses on the variation *within* the sub-metric EQRs. It assumes that the sub-metrics are measure different aspects of the biological community and that, even with perfect information, they would not necessarily give the same EQR result. In this situation, the uncertainty within the sub-metrics should be combined to give the uncertainty, and therefore CofC, of the final EQR.

A couple of hypothetical scenarios serve to illustrate the distinction between these approaches.

Scenario 1: Three sub-metrics (A-C) with different EQRs are each estimated without error (Table 2.1). The short-cut approach would say that the three sub-metrics are alternative measures of the same thing, and that because they give different results there must be some uncertainty in the Final EQR. By contrast, the bottom-up approach would say that the Final EQR is by definition the average of those three particular sub-metrics, and that if each sub-metric is known without error, then so must the Final EQR.

Table 2.1 Scenario 1

Sub-metric	Sub-metric EQR	SE of EQR
A	0.6	0.0
B	0.7	0.0
C	0.8	0.0

Scenario 2: Three sub-metrics (A-C) with equal EQRs, are imperfectly estimated (Table 2.2). The short-cut approach would say that because the sub-metrics give an identical answer, the Final EQR must be known without error. By contrast, the bottom-up approach would say that because each sub-metric has some uncertainty associated with it one would expect to get a different set of sub-metric EQR values if the sampling were repeated. For this reason, the Final EQR is not known without error.

Table 2.2 Scenario 2

Sub-metric	Sub-metric EQR	SE of EQR
A	0.7	0.2
B	0.7	0.1
C	0.7	0.3

The choice of approach boils down to whether the sub-metrics are random, replicate measures of the same aspect of the biological community, or unique, fixed metrics that measure different aspects of the community. In this project, the Agency and WRc agreed to adopt a bottom-up approach.

So what does this mean for the three biological tools under consideration in this project? The distinction between the two approaches is not relevant to the rocky shore tool because the sub-metrics are combined to give an overall EQR result for each survey, and it is the variability between the survey EQRs that determines the confidence of class for the waterbody. The same applies to the opportunistic macroalgae tool; however, when computing confidence of class for each individual survey, the bottom-up approach is used to calculate the uncertainty in the survey EQR. The issue of combining sub-metrics is most critical to the phytoplankton tool because the three sub-metrics are calculated at the waterbody level, and then combined to give a final EQR; again, the bottom-up approach is preferred.

2.4 General assumptions

All three confidence of class tools assume that surveys of the quality elements are conducted in such as way as to give a representative and unbiased measure of biological conditions across the whole waterbody throughout the whole reporting period. Statistical manipulation of the resulting data cannot compensate for poorly planned and executed field sampling; for example, if three surveys of a rocky shore were all conducted in the first year of a six year period, then there is no way of knowing whether the conditions observed in that year accurately reflect the conditions in the following five years. External estimates (from other waterbodies, or other time periods) of the expected variation from year to year can help to estimate the uncertainty in the results, but this is no substitute for a sampling scheme that measures directly the spatial and temporal variation in the target population.

3. CAPTAIN

3.1 Overview

CAPTAIN (Confidence And Precision Tool AIds aNalysis) calculates confidence of class for the WFD TraC Opportunistic Macroalgae Tool. It calculates the confidence of class for individual surveys (one at a time) and also assesses the confidence of class for the whole waterbody over the whole reporting period, using the EQR results from one or more surveys (multiple waterbodies simultaneously).

3.2 Background

The opportunistic macroalgae tool measures the extent and biomass of opportunistic macroalgae in inter-tidal habitats. One or more surveys are undertaken during each reporting period, with each survey covering the whole waterbody. Each survey yields an EQR between 0 and 1 and the status of the waterbody is evaluated as the mean EQR of all surveys conducted during the reporting period (Figure 3.1).

The EQR result from each survey is the average of five sub-metric EQRs. The sub-metrics are:

1. % cover of AIH – The average % cover of algae in the available intertidal habitat.

2. Biomass (g/m^2) per m^2 AIH – The average biomass of algae per m^2 in the available intertidal habitat.

3. Presence of entrained algae – The % of quadrats where algae is seen to be growing deeper than 3cm into the underlying sediment indicating the likelihood of regeneration.

4. Total affected area (ha) – The total extent of the algal bloom, measured in hectares and based on the external perimeter of the bloom.

5. Biomass (g/m^2) per m^2 affected area – The average biomass of algae per meter squared over the affected area only.

Each survey is conducted using a stratified random sampling scheme, whereby the waterbody is divided into one or more patches with differing levels of macroalgae and two or more replicate quadrats are randomly positioned in each patch. (Areas of the waterbody with no algae effectively form a separate stratum or patch which is not surveyed and where %cover and biomass are assumed to be zero.)

When there is more than one patch, the sub-metrics 1, 2 and 4 are calculated by weighting the result for each patch by the area of each patch. The exception is sub-metric 3 (%Entrained algae), where the sub-metric is simply an unweighted average of the %entrainment in each patch.

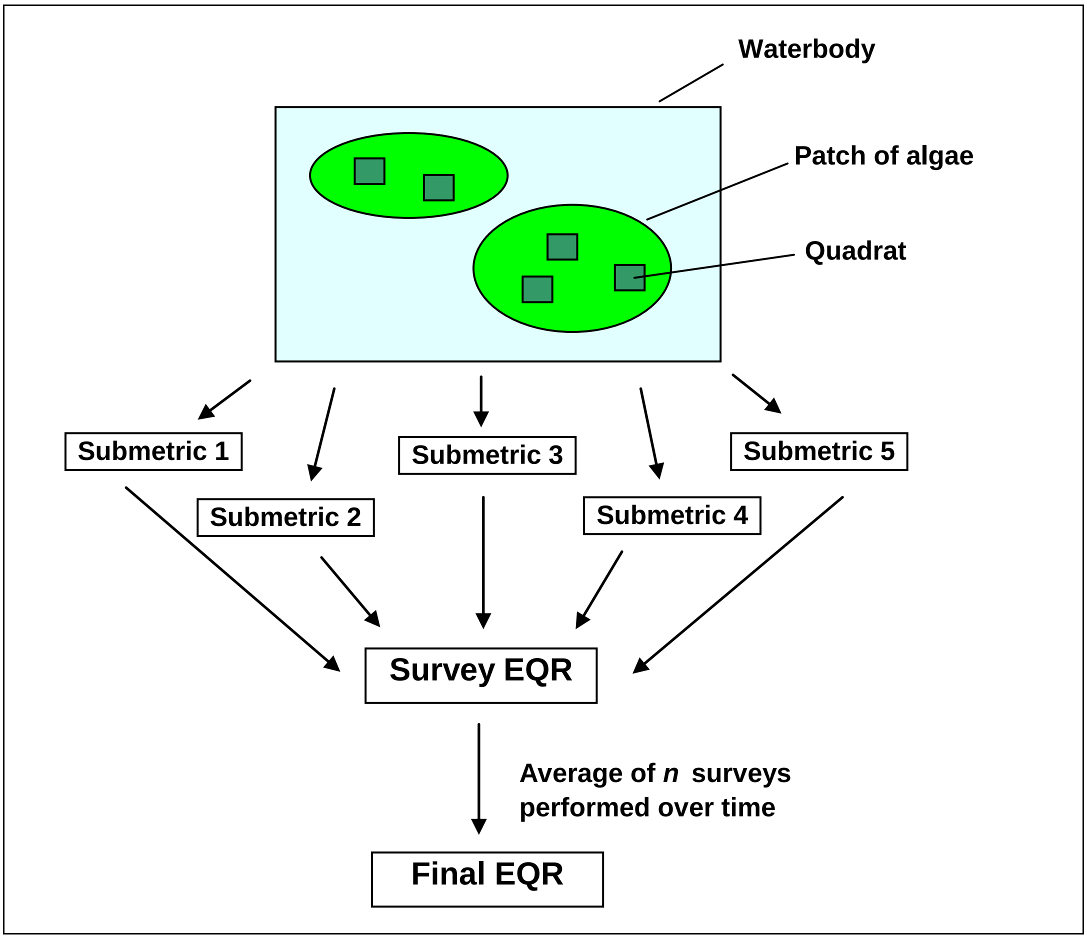

Figure 3.1 Sampling scheme for opportunistic macroalgae tool

3.3 Methodology

3.3.1 Introduction

CAPTAIN performs confidence of class calculations at three levels: sub-metric, survey and waterbody:

1. The confidence of class for each sub-metric in each survey is based on the sub-metric EQR and takes account of sampling error plus any error in the measurement of patch area.

2. The confidence of class for each survey is based on the Survey EQR and takes account of combined uncertainty in the five sub-metrics.

3. The confidence of class for the waterbody is based on the Final EQR and takes account of the temporal variation among the EQR results from replicate surveys.

This section first describes how each sub-metric score is derived, and its corresponding standard error is calculated to give a sub-metric CofC. Next, it considers how the sub-metric scores are combined to yield an EQR and CofC for each survey. Finally, the survey EQRs are combined to give a Final EQR and a CofC assessment is performed for the waterbody as a whole.

CAPTAIN adopts a flexible stance towards error in the measurement of patch areas and AIH. It is possible to assume that the area of each patch and the area of the AIH are measured without error. This is likely to be an unrealistic assumption, however, and so the tool allows the user to specify a relative standard deviation (RSD) to represent the likely uncertainty in these measurements. As an example, an RSD of 0.1 means that the standard deviation is 10% of the measurement; this equates to being 95% confident that that true area is within ± 20% of the measured area. It is suggested a default RSD value of 0.1 is used for both patch area and AIH unless more detailed information is available. It is, of course, possible to adjust the RSD values to see how sensitive the final confidence of class results are to the level of measurement error; if the results change very little as measurement error increases, then it is not worth worrying about getting a reliable estimate of the measurement error; on the other hand, if the results are very sensitive to the level of measurement error, then greater efforts should be made to obtain a reliable estimate of the measurement error.

3.3.2 Terminology

Table 3.1 defines the notation used to refer to the sampling results.

Table 3.1 Notation used by CAPTAIN

n	The number of surveys carried out in a waterbody during a reporting period
p_i	The number of patches sampled in the i^{th} survey
q_{ij}	The number of quadrats in the j^{th} patch in the i^{th} survey
x_{ijk}	The measurement taken in the k^{th} quadrat in the j^{th} patch in the i^{th} survey
s_{ij}^2	The variance of the k measurements in the j^{th} patch in the i^{th} survey
a_{ij}	The area of the j^{th} patch in the i^{th} survey
AIH_i	The Available Intertidal Habitat in the i^{th} survey
AA_i	The total Affected Area in the i^{th} survey = the sum of the areas of the p patches
Q_i	The sub-metric score in the i^{th} survey
Q_i^*	The sub-metric EQR in the i^{th} survey

3.3.3 CofC for sub-metric 1 (% cover of AIH)

Let x_{ijk} = the % cover of algae in the k^{th} quadrat. The average % cover in the j^{th} patch is the average of the % cover in the q quadrats:

$$\bar{x}_{ij} = \frac{\sum_{k=1}^{q} x_{ijk}}{q_{ij}} \quad (1)$$

The sub-metric score (Q_i) for the i^{th} survey is defined as the % cover of algae in the AIH and is calculated as a weighted average of the % cover in each of the p patches:

$$Q_i = \sum_{j=1}^{p} \left(\bar{x}_{ij} * \frac{a_{ij}}{AIH_i} \right) \quad (2)$$

When the patch areas and the AIH are measured without error, the only uncertainty is in the estimate of $\bar{x}_{ij}$, so the standard error of Q_i is given by:

$$SE(Q_i) = \sqrt{\sum_{j=1}^{p} \left(\frac{s_{ij}^2}{q_{ij}} \times \left(\frac{a_{ij}}{AIH_i} \right)^2 \right)} \quad (3)$$

where s_{ij}^2 is the variance in % cover of algae within each patch:

$$s_{ij}^2 = \frac{\sum_{k=1}^{q} (x_{ijk} - \bar{x}_{ij})}{q_{ij}} \quad (4)$$

When the patch areas (a_{ij}) and the AIH are subject to measurement error, the standard error calculations are more complicated. Let:

> $RSD(a_{ij})$ = the relative standard deviation in measurements of each patch area (assumed to be the same for all patches in each survey); and
>
> $RSD(AIH)$ = the relative standard deviation in measurement of the AIH.

To ease the calculations, CAPTAIN uses a re-arranged version of equation 4:

$$Q_i = \sum_{j=1}^{p} \left(\bar{x}_{ij} * \frac{a_{ij}}{AIH_i} \right) = \sum_{j=1}^{p} \frac{\bar{x}_{ij} * a_{ij}}{AIH_i} = \frac{\sum_{j=1}^{p} (\bar{x}_{ij} * a_{ij})}{AIH_i} = \frac{\sum_{j=1}^{p} X_{ij}}{AIH_i} = \frac{X_i}{AIH_i} \quad (5)$$

where X_{ij} represents the total area of algae in the j^{th} patch and X_i represents the total area of algae in all patches (i.e. in the affected area). (Note that, for convenience, $\bar{x}_{ij}$ is actually a proportion rather than a percentage in this context).

The relative standard deviation of X_{ij} is:

$$RSD(X_{ij}) = \sqrt{RSD(\bar{x}_{ij})^2 + RSD(a_{ij})^2} \quad (6)$$

and so the standard deviation of X_{ij} is:

$$SD(X_{ij}) = RSD(X_{ij}) * X_{ij} \quad (7)$$

Summing across the p patches, the standard deviation of X_i is:

$$SD(X_i) = \sqrt{\sum_{j=1}^{p} SD(X_{ij})^2} \quad (8)$$

Dividing by the AIH, the relative standard error of Q_i is:

$$RSE(Q_i) = \sqrt{RSD(X_i)^2 + RSD(AIH_i)^2} = \sqrt{\left(\frac{SD(X_i)}{X_i}\right)^2 + RSD(AIH_i)^2} \quad (9)$$

and the standard error of Q_i is:

$$SE(Q_i) = RSE(Q_i) * Q_i \quad (10)$$

Q_i and its standard error are converted into a confidence of class following the Normal distribution approach described in Appendix B.

3.3.4 CofC for sub-metric 2 (biomass per m^2 AIH)

Calculations are exactly as for sub-metric 1, except, x_{ijk} = the biomass of algae in the k^{th} quadrat and $\bar{x}_{ij}$ = the average biomass of algae in the j^{th} patch.

3.3.5 CofC for sub-metric 3 (presence of entrained algae)

If x_{ijk} = the presence or absence (as 1 or 0) of entrained algae in k^{th} quadrat, then the average of the % entrainment in the j^{th} patch is:

$$\bar{x}_{ij} = 100 * \frac{\sum_{k=1}^{q} x_{ijk}}{q_{ij}}$$

To then calculate the sub-metric score, standard deviation and standard error follow the calculations outlined in sub-metric 5, in section 3.3.7.

3.3.6 CofC for sub-metric 4 (total affected area)

The affected area at the time of the i^{th} survey (AA_i) is the combined area of the p patches that contain algae:

$$Q_i = AA_i = \sum_{j=1}^{p} a_{ij} \qquad (13)$$

When each patch area is measured without error, the standard error of Q_i will be zero.

When the patch areas are each subject to the same measurement error, quantified by $RSD(a_{ij})$, the standard error of Q_i will be:

$$SE(Q_i) = \sqrt{\sum_{j=1}^{p} SD(a_{ij})^2} = \sqrt{\sum_{j=1}^{p} \left[RSD(a_{ij}) * a_{ij}\right]^2} = \sqrt{RSD(a_{ij})^2 * \sum_{j=1}^{p} a_{ij}^2} \qquad (14)$$

Q_i and its standard error are converted into a confidence of class following the Normal distribution approach described in Appendix B.

3.3.7 CofC for sub-metric 5 (biomass per m^2 affected area)

The calculations follow those for sub-metric 1, except that AIH is replaced by the affected area (AA).

If x_{ijk} = the biomass of algae in the k^{th} quadrat, then the average biomass in the j^{th} patch is:

$$\bar{x}_{ij} = \frac{\sum_{k=1}^{q} x_{ijk}}{q_{ij}} \qquad (15)$$

and the sub-metric score (Q_i, the average biomass density in the AA) is:

$$Q_i = \sum_{j=1}^{p} \left(\bar{x}_{ij} * \frac{a_{ij}}{AA_i} \right) \qquad (16)$$

As before, when the patch areas, and hence the AA, are measured without error, the only uncertainty is in the estimate of $\bar{x}_{ij}$, so the standard error of Q_i is given by:

$$SE(Q_i) = \sqrt{\sum_{j=1}^{p}\left(\frac{s_{ij}^2}{q_{ij}} \times \left(\frac{a_{ij}}{AA_i}\right)^2\right)} \qquad (17)$$

When the AA is subject to measurement error, the standard error calculations are more complicated. As with sub-metric 1, $RSD(a_{ij})$ = the relative standard deviation in measurements of each patch area (assumed to be the same for all patches in each survey), and Q_i can be expressed as:

$$Q_i = \sum_{j=1}^{p}\left(\bar{x}_{ij} * \frac{a_{ij}}{AA_i}\right) = \sum_{j=1}^{p}\frac{\bar{x}_{ij} * a_{ij}}{AA_i} = \frac{\sum_{j=1}^{p}(\bar{x}_{ij} * a_{ij})}{AA_i} = \frac{\sum_{j=1}^{p} X_{ij}}{AA_i} = \frac{X_i}{AA_i} \qquad (18)$$

where X_{ij} represents the total biomass of algae in each patch and X_i represents the total biomass of algae in the affected area.

The relative standard deviation of X_{ij}, the total biomass in each patch, is:

$$RSD(X_{ij}) = \sqrt{RSD(\bar{x}_{ij})^2 + RSD(a_{ij})^2} \qquad (19)$$

and the standard deviation of X_{ij} is:

$$SD(X_{ij}) = RSD(X_{ij}) * X_{ij} \qquad (20)$$

Summing across the p patches, the standard deviation of X_i, the total biomass in the affected area, is:

$$SD(X_i) = \sqrt{\sum_{j=1}^{p} SD(X_{ij})^2} \qquad (21)$$

Dividing by the AA, the relative standard error of Q_i is:

$$RSE(Q_i) = \sqrt{RSD(X_i)^2 + RSD(AA_i)^2} = \sqrt{\left(\frac{SD(X_i)}{X_i}\right)^2 + \left(\frac{SD(AA_i)}{AA_i}\right)^2} \qquad (22)$$

and the standard error of Q_i is:

$$SE(Q_i) = RSE(Q_i) * Q_i \qquad (23)$$

Q_i and its standard error are converted into a confidence of class following the Normal distribution approach described in Appendix B.

3.3.8 CofC for Survey EQR

Each of the five sub-metric scores (Q_i) is normalised to give a sub-metric EQR between 0 and 1 (Q_i^*).

The Survey EQR ($\overline{Q}_i^*$) is computed as the average of the five sub-metric EQRs:

$$\overline{Q}_i^* = \frac{\sum Q_i^*}{5} \tag{24}$$

The standard errors of the sub-metric scores are converted to standard errors on the normalised EQR scale following the process described in Appendix C.

The standard errors of the five sub-metrics cannot be assumed to be independent because they are based on data from the same quadrats, and will share any errors in the measurement of patch area and AIH. The standard error of $\overline{Q}_i^*$ is therefore computed as:

$$SE(\overline{Q}_i^*) = \frac{\sum SE(Q_i^*)}{5} \tag{25}$$

$\overline{Q}_i$ and its standard error are converted into a confidence of class following the Normal distribution approach described in Appendix B.

3.3.9 CofC for Final EQR

Let $\overline{Q}_1^*$, $\overline{Q}_2^*$, …, $\overline{Q}_n^*$ represent a series of Survey EQRs derived from n surveys undertaken during the reporting period.

The Final EQR ($\overline{\overline{Q}}^*$) is given by:

$$\overline{\overline{Q}}^* = \frac{\sum_{i=1}^{n} \overline{Q}_i^*}{n} \tag{26}$$

The standard error of the Final EQR ($SE(\overline{\overline{Q}}^*)$) measures the uncertainty in the final status assessment and is given by:

$$SE(\overline{\overline{Q}}^*) = \frac{SD(\overline{Q}_i^*)}{\sqrt{n}} \tag{27}$$

where $SD(\overline{Q}_i^*)$ = the observed standard deviation of the Survey EQR values. Where only one survey is undertaken, the standard deviation is instead estimated from the mean EQR using the approach described in Appendix A.

The Final EQR and its standard error are converted to a confidence of class following the t-distribution approach set out in Appendix B.

4. PUGWASH

4.1 Overview

PUGWASH (Phytoplankton Uncertainty Gets Worked out And Statistically Handled) calculates confidence of class for the WFD TraC Phytoplankton Tool. It performs calculations for multiple waterbodies simultaneously and gives the confidence of class over the whole reporting period.

4.2 Background

The phytoplankton tool comprises three sub-metrics.

1. Elevated counts sub-metric - this is the average of four attributes, which each measure the proportion of sampling occasions on which a particular threshold level is exceeded. Those attributes are: phytoplankton biomass (mean chlorophyll), counts of any single species, counts of *Phaeocystis*, and counts of the total taxa.

2. Seasonal succession sub-metric - calculates the proportion of months in which community composition (as measured by a z-score) falls within a reference envelop for two major functional groups: diatoms and dinoflagellates. The sub-metric is the average of these two proportions.

3. Chlorophyll 90th percentile sub-metric - the 90th percentile of all chlorophyll concentrations during the growing season (March to September inclusive) is taken as a measure of phytoplankton biomass.

Each sub-metric is computed using data for the waterbody as a whole over a six year reporting period. Each sub-metric score is converted into an EQR via a two-step normalisation process. The first step converts the sub-metric score to an EQR scale between 0 and 1, where the status class boundaries are not equidistant (for example, Bad = 0.0 – 0.27, Poor = 0.27 – 0.34, Moderate = 0.34 – 0.44 etc). The second step transforms these EQR values onto an equal-width class scale (Bad = 0.0 – 0.20, Poor = 0.20 – 0.40, Moderate = 0.40 – 0.60 etc). For simplicity, PUGWASH combines these two normalisation steps into one, as illustrated in Appendix C.

The three sub-metric EQRs are then averaged to give a Final EQR between 0 and 1 (Figure 4.1).

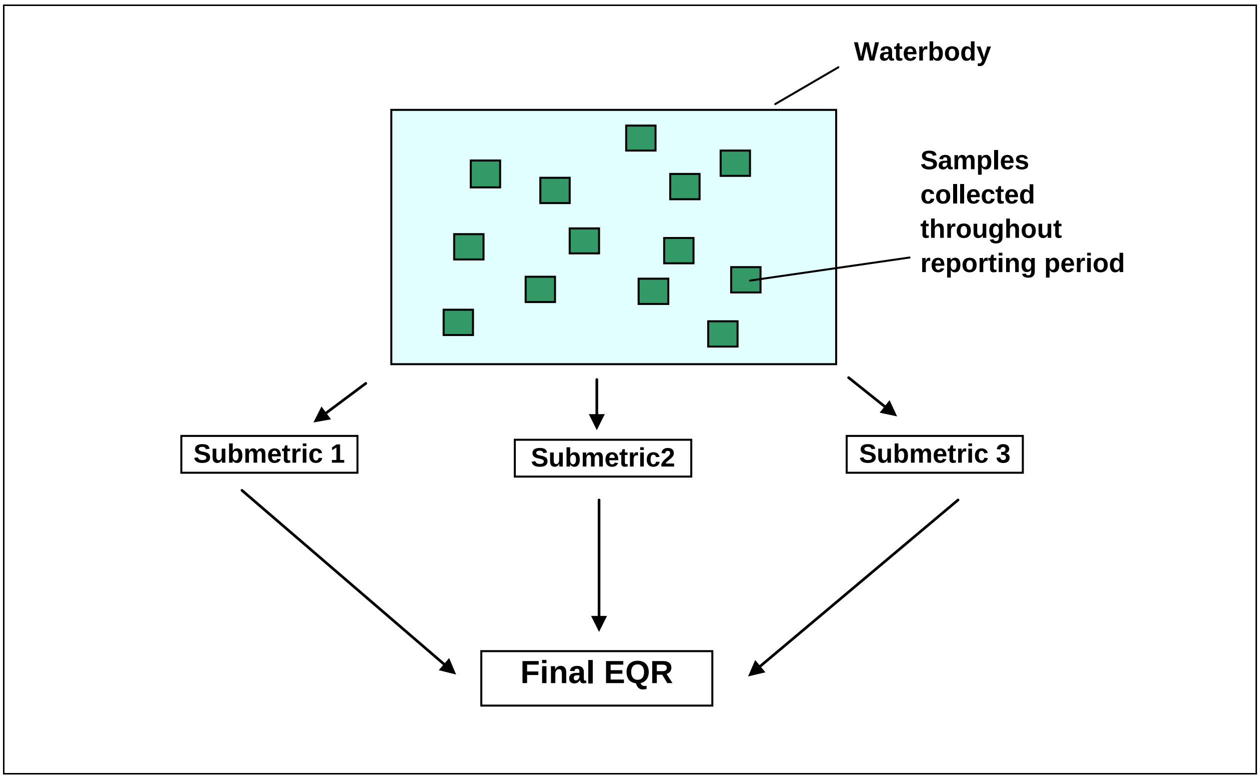

Figure 4.1 Sampling scheme for phytoplankton tool

4.3 Methodology

4.3.1 Introduction

As each sub-metric integrates spatial and temporal variability in the phytoplankton community, the uncertainty in the Final EQR is estimated by combining estimates of the uncertainty within each sub-metric EQR.

Briefly, PUGWASH adopts a bottom-up approach whereby each sub-metric score and its corresponding standard error are first used to compute the confidence of class for each sub-metric. Next, the three sub-metric scores are normalised to produce sub-metric EQRs between 0 and 1. Finally, the sub-metric EQRs are combined to give a Final EQR, and their standard errors are also combined to produce a confidence of class for the Final EQR result.

4.3.2 CofC for Elevated Counts sub-metric

Let n_1, n_2, …, n_k represent the number of samples taken of each of the $k = 4$ attributes (phytoplankton biomass, counts of any single species, counts of *Phaeocystis*, and counts of the total taxa) during the reporting period, and let r_1, r_2,…, r_k represent the corresponding number of samples that exceeded the specified threshold values.

The proportion of exceedances for the i^{th} attribute is given by:

$$p_i = \frac{r_i}{n_i} \tag{28}$$

The sub-metric score (Q) is calculated as the mean proportion of exceedances across the k attributes:

$$Q = \overline{p} = \frac{\sum_{i=1}^{k} p_i}{k} \tag{29}$$

Two alternative approaches were considered for calculating the standard error of Q. The original approach used a Normal approximation to estimate the standard error of each attribute proportion:

$$SE(p_i) = \frac{p_i(1-p_i)}{n_i} = \frac{r_i(n_i - r_i)}{n_i^3} \tag{30}$$

This formulation has the benefit that the standard errors of the four attributes can be easily combined to give the standard error of $\overline{p}$, but it gives unreliable results when the sample sizes are less than 30, and when the proportions are close to 0 or 1. Specifically, it was found that an attribute with very few samples and no exceedances would yield $p_i = 0$ and $SE(p_i) = 0$, giving the impression of no uncertainty when the paucity of replicate samples means that the uncertainty is actually very high.

As a result of these problems, PUGWASH utilises a less direct but more accurate method for calculating the standard error of Q, which proceeds as follows. First, a 95% confidence interval is constructed around each attribute proportion using the Wilson Score approach:

$$CI(p_i) = \frac{p_i + \frac{1}{2n_i} z^2 \pm z\sqrt{\frac{p_i(1-p_i)}{n_i} + \frac{z^2}{4n_i^2}}}{1 + \frac{z^2}{n_i}} \tag{31}$$

where z is the 97.5th percentile of a standard normal distribution, and takes a value of 1.96.

Second, the upper (UCL) and lower (LCL) 95% confidence limits are converted into an approximate standard error:

$$SE(p_i) \approx \frac{UCL(p_i) - LCL(p_i)}{2 \times 1.96} \tag{32}$$

Finally, the standard errors for each of the k attributes are combined to give a standard error for Q:

$$SE(Q) = \sqrt{\frac{\sum SE(p_i)^2}{k^2}} \quad (33)$$

The sub-metric score Q and its standard error are converted to a sub-metric confidence of class following the Normal distribution approach set out in Appendix B.

4.3.3 CofC for Seasonal Succession sub-metric

As the Seasonal Succession metric (Q) is the average proportion of two attributes (diatoms and dinoflagellates), the standard error of Q is computed in exactly the same way as for the Elevated Counts sub-metric (with $k = 2$ attributes).

4.3.4 CofC for Chlorophyll-a 90th Percentile sub-metric

The chlorophyll 90th percentile sub-metric (Q) is evaluated as the 90th percentile of all chlorophyll concentrations during the growing season (March to September inclusive).

PUGWASH estimates the 90th percentile non-parametrically as the mth ordered value, where:

$$m = n \times \frac{90}{100} \quad (34)$$

and n equals the number of chlorophyll-a samples. So for $n = 50$ samples, the 90th percentile is given by the 45th smallest concentration (i.e. the 5 largest). A minimum of nine samples are required to compute the 90th percentile.

(Note: this formulation is preferred to the more conventional Weibull approach, which gives $m = (n-1)*(90/100)$. The Weibull approach assumes (correctly) that the chlorophyll-a measurements are a sample from a wider population, but it occasionally produces a contradiction between the face-value class and the most probable class in the CofC assessment because of the way that the CofC calculations work (see below). It is important to be aware that the use of equation 34 changes subtly the definition of the chlorophyll sub-metric: strictly, a waterbody has to have a 90th percentile ≤ a threshold to be in the higher status class (i.e. 90% of the *population* of possible measurements have to be ≤ the threshold), whereas PUGWASH assesses whether a waterbody has 90% of *sampled measurements* ≤ the threshold. In most cases, this makes no appreciable difference to the results.)

Q is converted to a confidence of class using a binomial model. Let the four intermediate sub-metric score boundaries be denoted by L_5, L_4, L_3 and L_2 (in the order Bad/Poor → Good/High). The aim is to determine the levels of confidence we have that the true quality is respectively in Class 5, 4, 3, 2 and 1. To do this, we first do four calculations. For each class boundary 'i' in turn, we determine the number of samples (r_i) that fall below L_i, and ask: What is the probability p_i of observing r_i or fewer out of n samples if the probability of each individual sample being below L_i is exactly 0.9 (i.e. the 90th percentile is on the L_i boundary)?

p_i is calculated using the cumulative probability function of a binomial distribution with $n = n$ and $p = 0.9$.

This enables us to make the following five statements:

- Confidence of class 5 (Bad) = $100(1-p_5)$.
- Confidence of class 4 (Poor) = $100(p_5 – p_4)$.
- Confidence of class 3 (Moderate) = $100(p_4 – p_3)$.
- Confidence of class 2 (Good) = $100(p_3 – p_2)$.
- Confidence of class 1 (High) = $100p_2$.

Note that these five quantities sum to 100%.

The standard error of Q is approximated by first constructing a 90% confidence interval around the percentile estimate. The lower confidence limit is taken to be the q^{th} smallest value in the dataset, where: q is the largest integer for which the cumulative binomial distribution (*n* = *n*; *p* = 0.9) is ≤ 0.05. Similarly, the upper confidence limit is taken to be the r^{th} smallest value in the dataset, where r is the smallest integer for which the cumulative binomial distribution (*n* = *n*; *p* = 0.9) is ≥ 0.95.

For example, if *n* = 49 samples, the lower 90% confidence limit is given by the 39th smallest value (i.e. 11th largest) because the probability of getting 39 or fewer ‘successes’ out of 49 when *p* = 0.9 is 0.021. Similarly, the upper 90% confidence limit is given by the 47th smallest value (i.e. 3rd largest) because the probability of getting 47 or fewer ‘successes’ out of 49 when *p* = 0.9 is 0.963.

The upper (UCL) and lower (LCL) 90% confidence limits are converted into an approximate standard error:

$$SE(Q) = \frac{UCL(Q) - LCL(Q)}{2 \times 1.65} \tag{35}$$

4.3.5 CofC for Final EQR

Before computing the Final EQR, the three sub-metric scores Q and their standard errors are normalised onto an equal class-width scale running from 0 to 1 as described in Appendix C.

The $a = 3$ sub-metric EQRs (Q^*) are averaged to give a Final EQR between 0 and 1 ($\overline{Q}^*$).

The standard error of the Final EQR is computed from the standard errors of the sub-metric EQRs, assuming that the three sub-metrics are independent:

$$SE(\overline{Q}^{*}) = \sqrt{\frac{\sum_{i=1}^{a} SE(Q^{*})^{2}}{a^{2}}} \qquad (36)$$

The Final EQR and its standard error are then converted to a confidence of class following the Normal distribution approach set out in Appendix B.

5. PIRATES

5.1 Overview

PIRATES (Precision In Rocky shores Analysed To Extract Statistics) calculates confidence of class for the WFD TraC Reduced Species List (RSL) Tool. It performs calculations for multiple waterbodies simultaneously and gives the confidence of class over the whole reporting period.

5.2 Background

Surveys of macroalgal community composition are conducted on one or more rocky shores in each waterbody, on one or more occasions during each reporting period. Each survey yields five sub-metric EQRs, which are averaged to give a Survey EQR between 0 and 1. Status is defined by a Final EQR, which is the mean of the Survey EQR values (Figure 5.1).

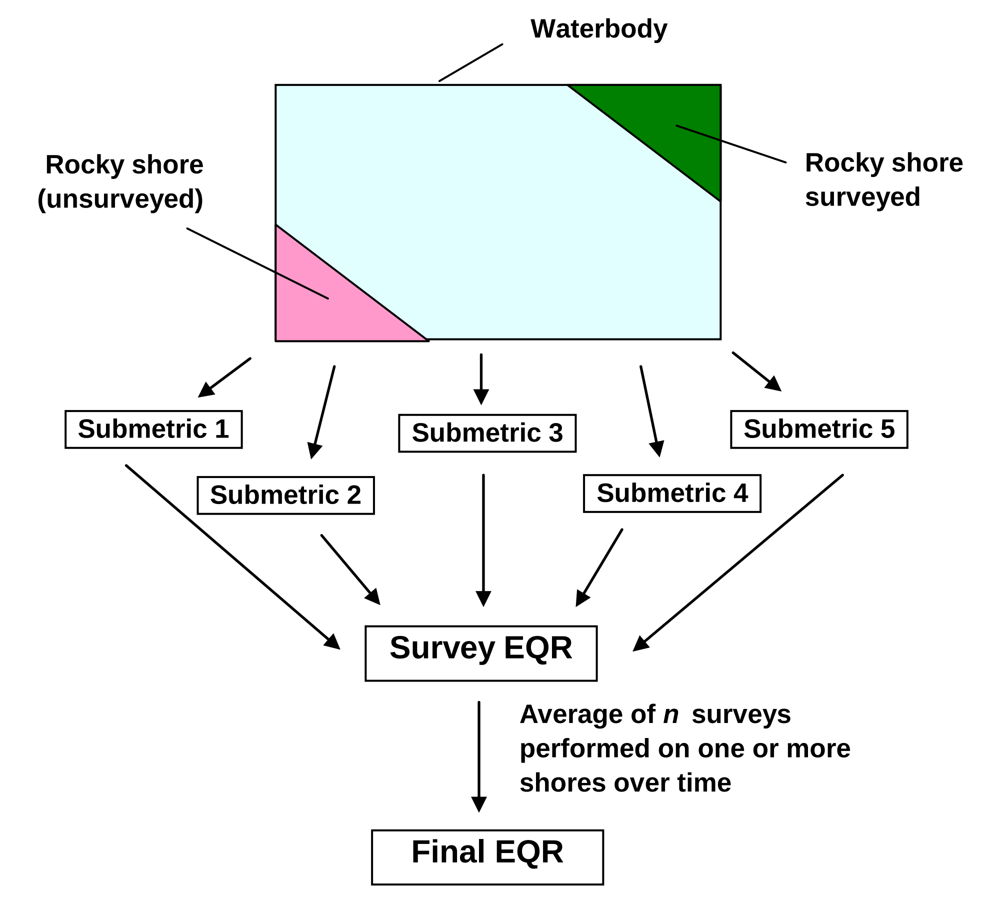

Figure 5.1 Sampling scheme for RSL tool

The uncertainty in the Final EQR is estimated by quantifying the spatio-temporal variation among the Survey EQR results.

5.3 Methodology

Let $\overline{Q}_1^*$, $\overline{Q}_2^*$, …, $\overline{Q}_n^*$ represent a series of Survey EQRs derived from n surveys undertaken during the reporting period.

The Final EQR ($\overline{\overline{Q}}^*$) is given by:

$$\overline{\overline{Q}}^* = \frac{\sum_{i=1}^{n} \overline{Q}_i^*}{n} \tag{37}$$

The standard error of the Final EQR ($SE(\overline{\overline{Q}}^*)$) measures the uncertainty in the final status assessment and is given by:

$$SE(\overline{\overline{Q}}^*) = \frac{SD(\overline{Q}_i^*)}{\sqrt{n}} \tag{38}$$

where $SD(\overline{Q}_i^*)$ = the observed standard deviation of the Survey EQR values. Where only one survey is undertaken, the standard deviation is instead estimated from the mean EQR using the approach described in Appendix A.

The Final EQR and its standard error are converted to a confidence of class following the t-distribution approach set out in Appendix B.

APPENDIX A

To calculate confidence of class for the Final EQR generated by CAPTAIN and PIRATES it is vital to have an estimate of the uncertainty in the Final EQR. This uncertainty arises from within-waterbody spatial and temporal variability in the biological community, which can be measured only by having two or more replicate surveys conducted in the waterbody during the reporting period. Where a waterbody has just a single independent survey and associated EQR result, this variability cannot be measured directly, but can be estimated indirectly using data from other waterbodies. This Appendix describes an approach developed by Ellis & Adriaenssens (2006) to estimate the likely spatio-temporal variability in Survey EQR as a function of the mean Survey EQR in a waterbody.

The approach seeks to model the combined spatial and temporal variability in Survey EQR results, as measured by their standard deviation, as a function of the mean EQR in a waterbody. Variability is expected to greatest in waterbodies of moderate status (EQR ≈ 0.5), and to get progressively smaller as the mean EQR tends towards 0 or 1. If this does not seem intuitive, then consider that for a waterbody to have a mean EQR of exactly 0 (or 1), all surveys must yield EQR values of 0 (or 1) – i.e. there must be no variation among surveys.

A power curve is used to capture this ∩-shaped relationship. The curve takes the form:

$$SD(X_i) = a + b_1\overline{X}_i + b_2\overline{X}_i^k$$

where:

$SD(X_i)$ = the standard deviation of replicate EQR results in waterbody *i*.

$\overline{X}$ = the mean Survey EQR in waterbody *i*.

a = intercept (the standard deviation when the mean EQR = 0; this is usually fixed at 0.01 to anchor the curve and to ensure that the model always produces non-zero standard deviation values).

b_1 and b_2 are regression coefficients, and k is a power coefficient, that together define the shape of the curve.

Figure A1 illustrates a typical dataset with a power curve fitted to it. Each black dot represents one waterbody, the blue squares represent the anchor points at Mean EQR = 0 and Mean EQR = 1, and the red line represents the best-fit power curve.

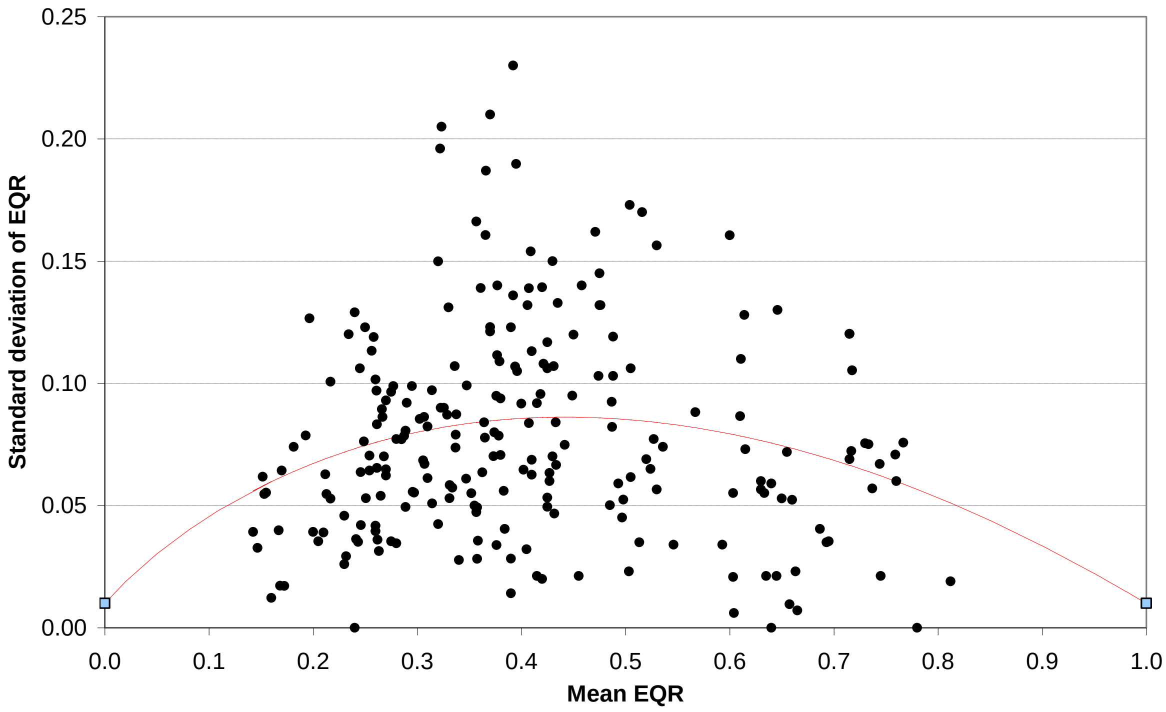

Figure A1 **A power curve describing the relationship between EQR variability and mean EQR**

Now, suppose that we have a waterbody with just a single survey, yielding an EQR of, say, 0.8. We take that value of 0.8 as the best available estimate of the mean EQR in that waterbody, and use the power curve to estimate the likely variability that we would have observed had we had two or more replicate EQR results. Using Figure A1, we estimate the standard deviation to be 0.053. With just a single EQR result (*n* = 1), the standard error of the mean EQR is therefore given by:

$$SE(\bar{X}) = \frac{SD(X)}{\sqrt{n}} = \frac{0.053}{\sqrt{1}} = 0.053.$$

APPENDIX B

After estimating the EQR and its associated uncertainty, it is necessary to decide on a suitable statistical model for the uncertainty in the EQR. The simplest option is to assume that the EQR uncertainty is Normally distributed around the specified true EQR value, with the predicted standard deviation. However, although this model is quite acceptable for most values of EQR it becomes unsatisfactory at either extreme, because the assumed Normal distribution 'spills' outside the permitted 0-1 range.

For this reason Ellis & Adriaenssens (2006) adopted the logit transformation, whereby the estimated EQR (Q^*) is transformed to a new variable Z given by:

$$Z = \ln\left(\frac{Q^*}{1-Q^*}\right)$$

where ln denotes 'logs to base e'. As Q^* runs from 0 to 1, the transformed variable Z runs from -∞ to +∞, and so there is no longer any risk of spillage. Thus it is possible to safely use the assumption of Normal error in the logit world, and then transform the resulting distribution back into the EQR world.

This is easier to see with the help of a diagram. Figure A1 shows the situation in which the assumed EQR mean and standard error are 0.85 and 0.10, respectively. Under the simple Normality assumption, an appreciable part of the right-hand tail spills beyond EQR = 1. In contrast, the logit transformation ensures that the error distribution ends asymptotically at 1 (at the expense of a longer left-hand tail so as to achieve the required standard deviation).

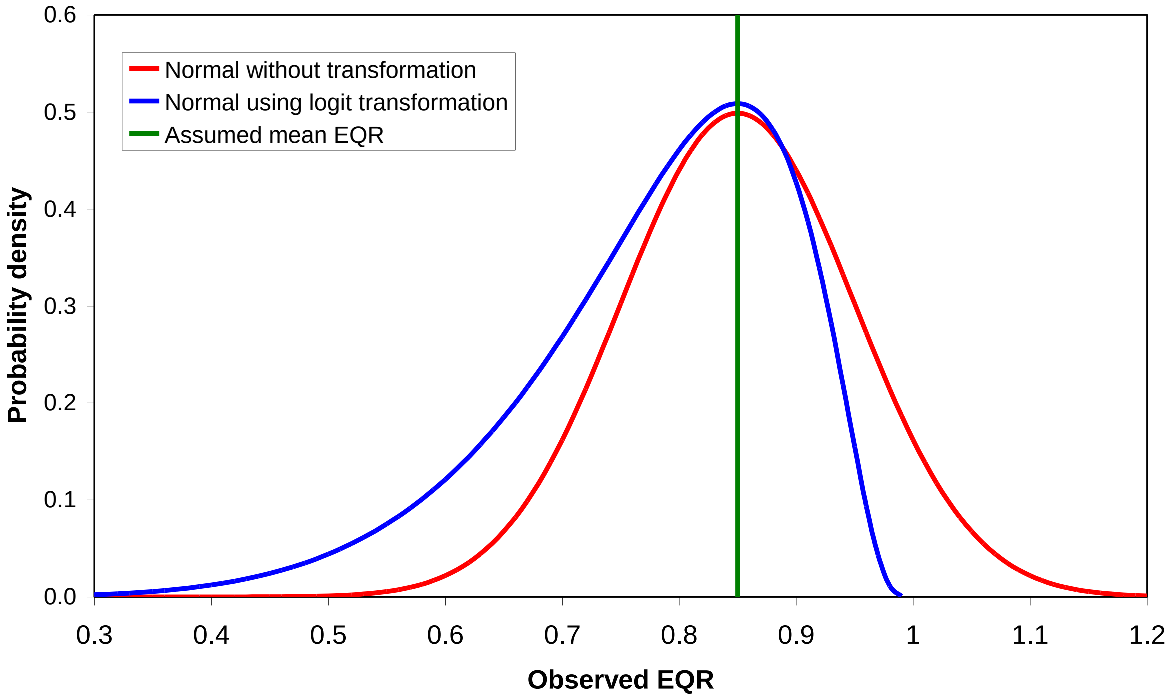

Figure B1 Illustration of the effect of the logit transformation of EQR

The confidence of class is then calculated in one of two ways.

If the EQRs that make up the Mean EQR result constitute the entire population (e.g. where the Mean EQR is the average of three sub-metric EQRs), then the confidence of class is computed using the Standard Normal Distribution. If, on the other hand, the EQRs that make up the Mean EQR result are just a random sample from a population of possible EQR results, then the confidence of class is computed using the *t*-distribution, which takes into account the additional sampling error.

Normal distribution approach

Let the four intermediate class boundaries be denoted by L_5, L_4, L_3 and L_2 (in the order Bad/Poor → Good/High). Suppose we observe an EQR value of Q^*, with a standard error of $SE(Q^*)$. The aim is to determine the levels of confidence we have that the true quality (at the time and place of sampling) is respectively in Class 5, 4, 3, 2 and 1. To do this, we first do four calculations. For each class boundary 'i' in turn, we ask the question: What is the probability p_i of observing an EQR of Q^* or better if the true mean quality, μ, were on the L_i boundary? This can be calculated as:

$$p_i = 1 - \Phi\left(\frac{Q^* - \mu}{SE(Q^*)}\right)$$

where Φ denotes the cumulative Normal probability.

We can turn this into a confidence statement by inverting it in the customary way, giving:

$$\text{Confidence}\,(Q^* \geq L_i) = 100(1 - p_i)$$

This enables us to make the following five statements:

- Confidence of class 5 (Bad) = $100p_5$.
- Confidence of class 4 (Poor) = $100(p_4 - p_5)$.
- Confidence of class 3 (Moderate) = $100(p_3 - p_4)$.
- Confidence of class 2 (Good) = $100(p_2 - p_3)$.
- Confidence of class 1 (High) = $100(1 - p_2)$.

Note that these five quantities sum to 100%.

T-distribution approach

The *t*-distribution approach works in exactly the same way as the Normal distribution approach, except that:

$$p_i = 1 - t\left(\frac{Q^* - \mu}{SE(Q^*)}\right)$$

where *t* denotes the cumulative *t* probability distribution with degrees of freedom = the number of individual EQR results contributing to the EQR result.

APPENDIX C

All three biological tools estimate one or more sub-metric scores, and then convert the sub-metric scores into an EQR via a two-step normalisation process. The first step converts the sub-metric score to an EQR scale between 0 and 1, where the status class boundaries are not equidistant (for example, Bad = 0.0 – 0.27, Poor = 0.27 – 0.34, Moderate = 0.34 – 0.44 etc). The second step transforms these EQR values onto an equal-width class scale (Bad = 0.0 – 0.20, Poor = 0.20 – 0.40, Moderate = 0.40 – 0.60 etc), as shown in Figure C1.

step 1 **step 2**

Class	Sub-metric score	Non-equidistant EQR	Equi-distant EQR
High	0--10	0.67 - 1.00	0.8 - 1.0
Good	10--15	0.44 - 0.67	0.6 - 0.8
Moderate	15--20	0.34 - 0.44	0.4 - 0.6
Poor	20--25	0.27 - 0.34	0.2 - 0.4
Bad	>25	0.00 - 0.27	0.0 - 0.2

Combined steps

EXAMPLES			
High	5.00	0.84	0.90
Good	13.00	0.53	0.68
Moderate	19.00	0.36	0.44
Poor	25.00	0.27	0.20
Bad	26.00	0.26	0.19

Figure C1 Normalisation of sub-metric scores to produce an EQR

The normalisation process creates problems when attempting to calculate the standard error associated with each sub-metric EQR. It is relatively straightforward to calculate the standard error of the sub-metric *score*, but there is no easy way to then normalise the standard error onto an equal-width EQR scale. The solution adopted in this study was to estimate a 95% confidence interval around each sub-metric score, normalise the upper and lower confidence limits, and then to derive an approximate standard error on the normalised EQR scale as:

$$SE(EQR) \approx \frac{UCL(EQR) - LCL(EQR)}{2 \times 1.96}$$

For example, using the normalisation procedure in Figure C1, if the sub-metric score was estimated to be 19, with a 95% confidence interval of 13 – 25, then this would translate into an EQR of 0.44 with a 95% confidence interval of 0.20 – 0.68. The standard error of the EQR would then be approximated as:

$$SE(EQR) \approx \frac{0.68 - 0.20}{2 \times 1.96} = 0.12$$