Collaborative Project (large-scale integrating project)
Grant Agreement 226273
Theme 6: Environment (including Climate Change)
Duration: March 1st, 2009 – February 29th, 2012

SEVENTH FRAMEWORK PROGRAMME

WISER

Water bodies in Europe: Integrative Systems to assess Ecological status and Recovery

DELIVERABLE

Deliverable D6.1-1: Report on a workshop to bring together experts experienced with tool development and uncertainty estimation

Lead contractor: **Centre for Ecology and Hydrology (CEH)**
Contributors: **Ralph Clarke & Iwan Jones**

Due date of deliverable: **Month 6**
Actual submission date: **Month 6**

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013) Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Content

Abbreviations:

WB = Water Body, the entity of water management according to the Water Framework Directive (WFD)

BQE = Biological Quality Element, organism groups demanded for assessment and monitoring of water ecological status (fish, benthic macroinvertebrates, aquatic macrophytes, angiosperms, marine macroalgae, benthic freshwater algae, phytoplankton)

EQR = Ecological Quality Ratio, ratio of observed assessment index value to the expected value under reference conditions; EQRs represent normalised index values on a numerical scale between 0 and 1

Rationale for workshop

Any indices of ecological quality and derived estimates of ecological status class of any water body (WB) with regard to the Water Framework Directive are of limited value without some understanding and quantitative estimates of their susceptibility to sampling variation and other sources of uncertainty. In recognition of this, the Water Framework Directive (WFD) (Council of the European Communities 2000) states "Estimates of the confidence and precision of the results provided by the monitoring programmes shall be given in the plan" (WFD Annex V, Section 1.3 "Monitoring of ecological status of surface waters") (Figure 1).

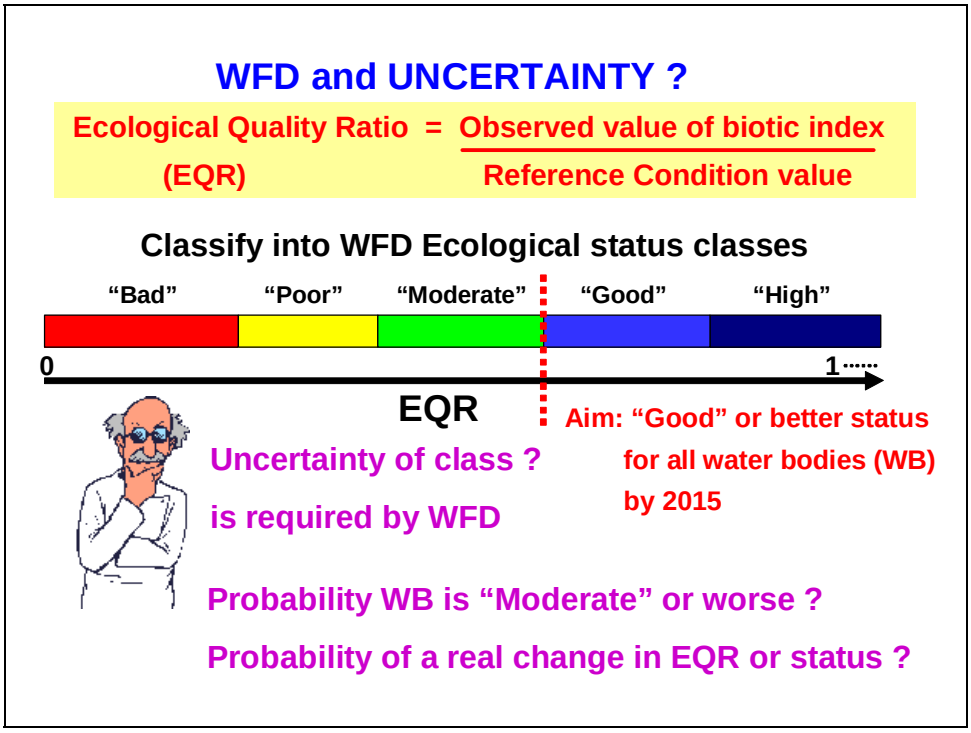

Figure 1 Understanding and estimates of uncertainty in assigning water bodies to WFD status class are requirements of the WFD.

The WISER project is primarily concerned with investigating and enhancing effective use of biological metrics and tools to assess the ecological status of lake (Module 3), coastal and transitional waters (Module 4). For lake WBs, the Biological Quality Elements (BQEs) to be considered and assessed are phytoplankton (WP3.1), macrophytes (WP3.2), macroinvertebrates (WP3.3) and fish (WP3.4). For coastal and transitional waters WB, the BQEs to be assessed for performance are phytoplankton (WP4.1), benthic macroflora (WP4.2), macroinvertebrates (WP4.3) and fish (WP4.4). Each of these WPs include objectives to assess and compare the uncertainty of individual metrics and/or methods arising from the various potential sources of sampling variation and errors. The relative precision of metrics will be used to help guide the final selection of metrics for assessing each WB type and estimates of sampling precision of selected metrics will be used to calculate estimates of uncertainty of estimates of ecological status class (WFD classes 'high', 'good', 'moderate', 'poor' and 'bad').

In addition, in Module 5, the objectives include developing statistical or mechanistic models relating the biota of WB to their physical characteristics and the driving pressure variables in order to predict the response to management and climate change within rivers (WP5.1), lakes (WP5.2) and coastal and transitional waters (WP5.3). These WP includes aims to assess the uncertainty associated with the use of the models to predict the change in pressure required to achieve or maintain good ecological status.

Thus the assessment of uncertainty due to natural spatial and temporal variation, sampling methodology, predictive modelling and other aspects of the bioassessment methodology are an important component of the WISER project.

To help all of the WISER partners understand and assess aspects of sampling variation and WB assessment uncertainty, the project includes a special Workpackage WP6.1 on uncertainty. It was important to ensure that assessments of the important components of sampling variation were included in each WP with planned field sampling programmes (Modules 3 and 4), most of which were to be carried out during the early stages of the project. Therefore a workshop on uncertainty was arranged for the second day of the week-long project start-up meeting of the whole project consortium held in Mallorca on 9-13 March 2009.

(Minutes from the full Kick-off meeting are available on the WISER Intranet as Deliverable D1.5.)

Workshop summary overview

The uncertainty workshop was held within the week-long project start-up meeting of the whole WISER project consortium held in Mallorca on 9-13 March 2009. Following general introduction, overview and administration presentations by Daniel Hering and others from the UDE project lead partner on the first afternoon, the uncertainty workshop was held throughout the next day to encourage maximum consideration of the major sources of variation (spatial, temporal and methodological) for each type of WB and BQE being assessed by each WP group.

The workshop was led by Ralph Clarke (Bournemouth University) and Iwan Jones (Centre for Ecology and Hydrology (CEH), now of Queen Mary University of London (QMUL)).

The first session involved "education and training" presentations by Ralph Clarke, Iwan Jones and Jacob Carstensen (National Environmental Research Institute (NERI)) who are all researchers experienced in the assessment of aquatic sampling and modelling uncertainty

Following these formal presentations to guide WISER partners, people then split up into small breakout groups for each WP in Modules 3 and 4 (i.e. for each WB type by BQE combination) to identify the sources of variation in their WB type (i.e. lake, coastal or transitional water) and in the methods used to sample that BQE.

Throughout the rest of the informal uncertainty workshop and continuing through the WISER start-up meeting, each Module continued discussions and made progress on agreeing the number of WB (termed 'sites') to be sampled/surveyed of each type and quality class, in which countries and by which partners. The aim was to sample/survey all, or as many BQE as possible, for each WB selected for the WISER field campaign.

One of the primary aims of the WISER field sampling campaign is to assess the relative size and importance of difference sources and scales of spatial and temporal variation on uncertainty in individual biological metrics and derived estimates of WB status class. Therefore discussions within WP groups, guided by the earlier formal presentations and with support from WP6.1 leaders Ralph Clarke and Iwan Jones, led to draft proposals for each BQE on the number of:

(i) stations/transects to be sampled in each WB (estimates spatial variability)

(ii) number of replicate samples to be taken at each station and by how many different people (estimates sampling method replicate and operator variability)

(iii) number of sub-samples or counts to be made from each sample and by how many different people (estimates sample sub-sampling, operator processing and identification variability)

It was agreed that, within the 3-year time and resource constraint of the project, the field campaign could not adequately assess temporal variability within WB and that any estimates of temporal variability in the biota and derived biological metrics would be obtained from existing data to be collated within WISER.

The initial field sampling programmes developed within the uncertainty workshop follow-up sessions within the Kick-off meeting have subsequently been revised to agree a final set of sites to be sampled/survey and by whom and with details on the precise sampling/surveying and sampling processing protocols and methods to be used for each BQE.

Workshop introduction with practical exercise

The presentation session was introduced by Ralph Clarke and Iwan Jones with a practical exercise ("ping-pong bingo") to illustrate the general issue of uncertainty and help bond WISER partners in the tasks ahead. Workpackage leaders were asked to take one sample each (an enumerated ping-pong ball) out of our water body (a bucket filled with water) to illustrate the variation between sample estimates of the true average value of the WB. In the illustrative small WB, the parameter we have chosen to use to represent the WB quality is the mean value of all possible samples from the WB; mean values of 40 or more are to treated as indicative that the WB is of good or better WFD status class, while WB with mean values less than 40 are used to indicate moderate or worse class. In our workshop WB, the true mean of the 'metric' was 42 – which was not only indicative of 'good' status but reputedly "the answer to everything" (Douglas Adams, Hitchhiker's Guide to the Galaxy, 1978) (Figure 2). However, in the WISER real world, we can only estimate the truth and try to minimise and quantify our uncertainty.

The various WISER WP leaders each took a random sample (ball) and the frequency distribution of the metric values was built-up with all values less than the high/good status class boundary value of 40 being classed as of "poor" or worse status. Figure 2 shows the true underlying frequency distribution of all possible sample values from our WB (bucket) whose mean value we are trying to estimate and indicate that using just one sample to estimate WB status, there was a 65% probability of classifying the WB correctly as 'good' or better and thus a 35% of mis-classifying the WB as of moderate or worse status and perhaps spending unnecessary resource on trying to develop and implement a management plan to improve the status of the WB.

The practical exercise continued by showing that if two (or more) rather than just one single (random) sample could be taken from a WB then the average of the metric values for the two (or more) samples would give a more precise estimate of the true mean value of the metric for this WB (Figure 3). Ralph Clarke then pointed out that, since

standard error of the sample mean = SE = SD/√n, where

SD = standard deviation between individual sample values and

n = number of samples,

then improvements in confidence of class can sometimes also be achieved by improvements in the basic sampling methodology which reduces the inherent variability (i.e. SD) between samples. This could by potentially be achieved by using different sampling equipment, sampling a larger area per sample, or sorting, identifying and counting a larger fraction of the field sample

back in the laboratory. Cost-effectiveness of different sampling techniques was discussed further by Iwan Jones (see below).

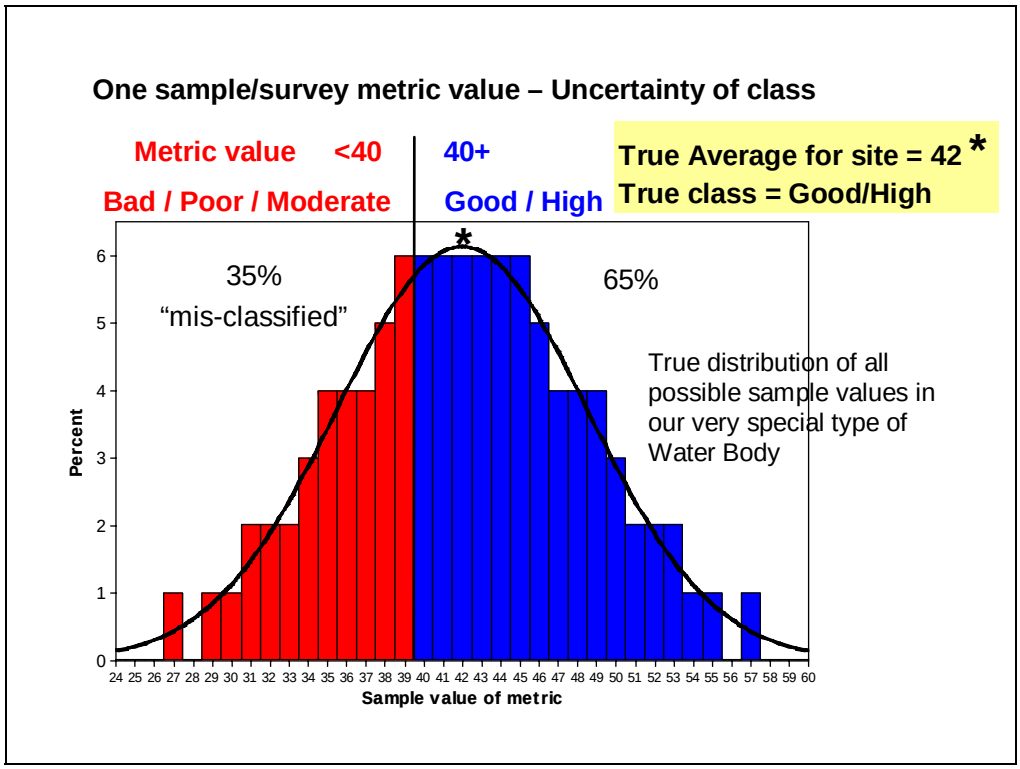

Figure 2 Frequency distribution of all possible samples in our special 'bucket' water body.

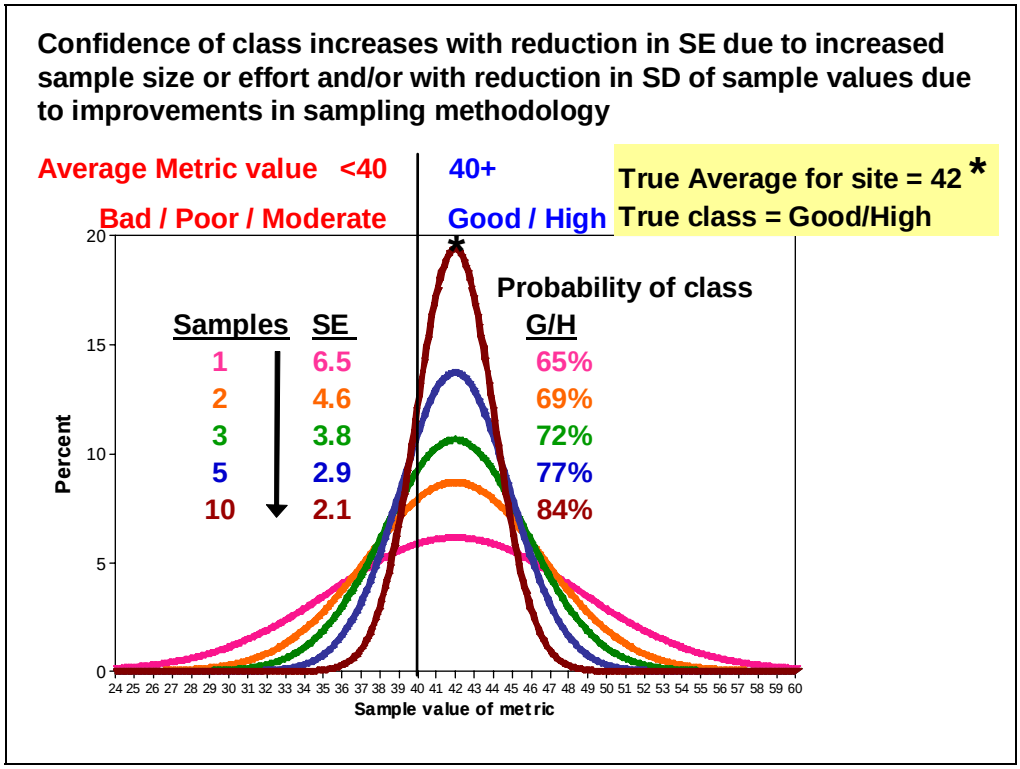

Figure 3 Demonstration of effect of taking two or more samples or otherwise reducing the standard error of the sampling method estimate on confidence of class.

Jannicke Moe (NIVA) pointed out that for single metrics and assuming a normal distribution of sampling variability, the NORMDIST function in EXCEL could be used to calculate confidence of class. In the workshop practical exercise example, if the true mean metric value is 42, the good/moderate boundary value is 40, and the SE is 6.5 (as with single samples in Figure 3), then the probability of correctly classifying the WB can be calculated in EXCEL by the function:

1-NORMDIST(40,42,6.5,TRUE).

Confidence of class and risk of mis-classification

Ralph Clarke showed how the confidence we have that a WB belongs to each particular status class depends on the value of the metric in relation to the metric's class boundaries and the sampling standard error of our estimate of the metric (Figure 4). In reality WB are often classified not on the basic of the observed values of biological metrics (such as taxonomic richness), but by the Ecological Quality Ratios (EQR) whereby the observed metric values have been standardised by the metric value expected for that type of WB if it was in 'Reference Condition' (Figure 1), as described later in the workshop for the UK RIVPACS (River Invertebrate Prediction And Classification System) (Wright et al. 2000).

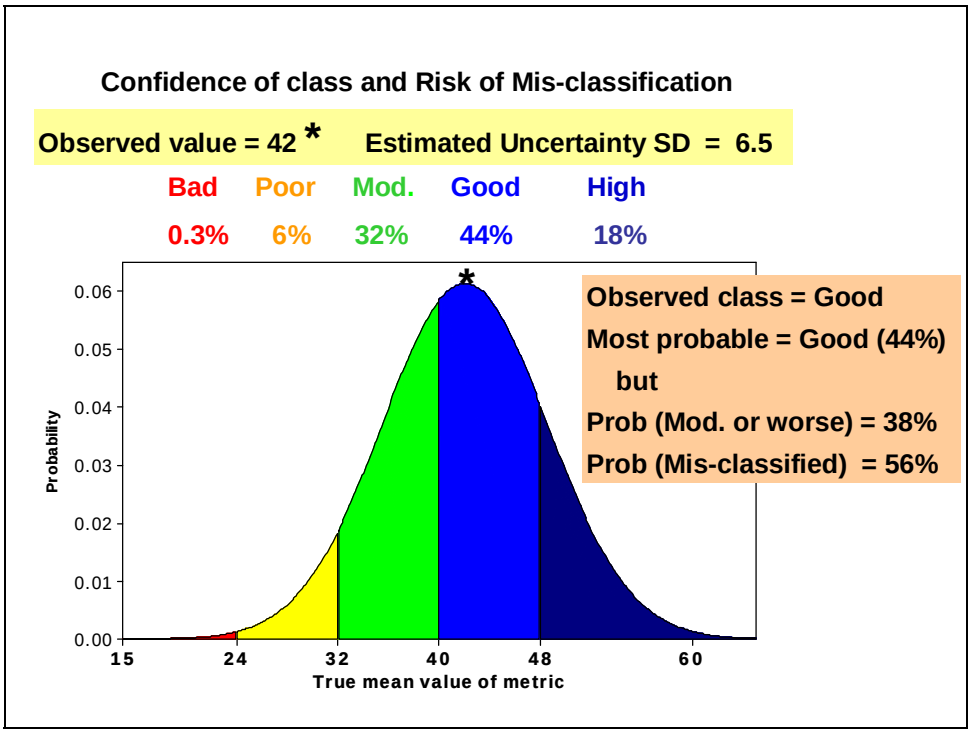

Figure 4 Illustration of confidence of class and risk of mis-classification.

Ralph Clarke highlighted how the confidence of class and the probability of misclassify a particular WB depends on the true mean EQR value in relation to the EQR class limits and can often be calculated mathematically by expressing the uncertainty SE as a percentage of the width of the status class intervals for the EQR (Figure 5).

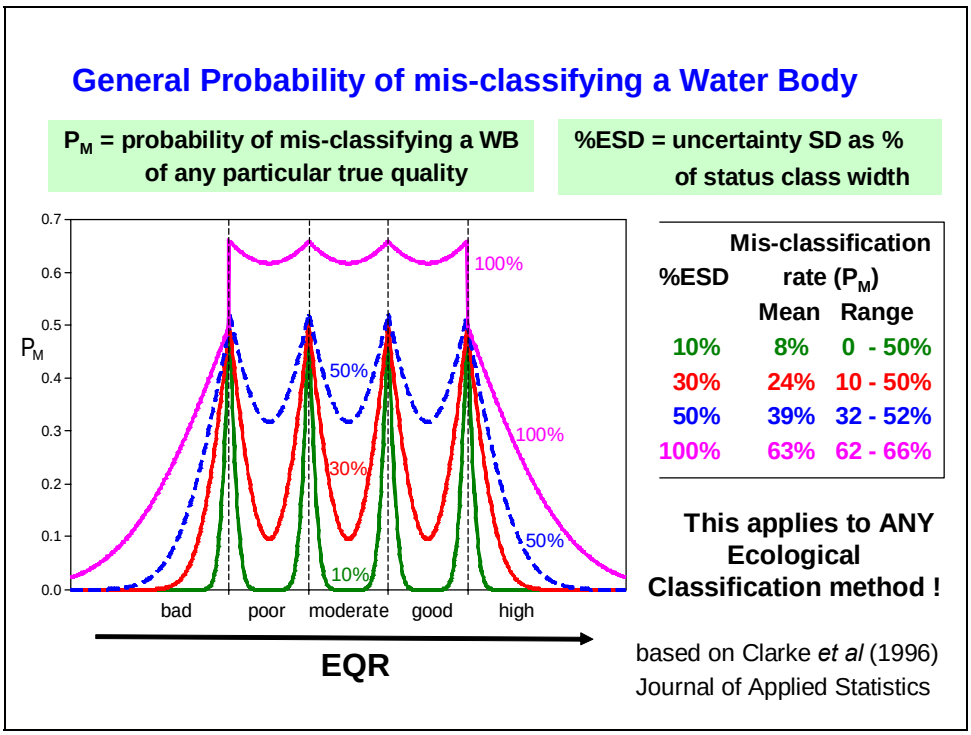

Figure 5 Illustration of probability of mis-classifying a water body based on a single EQR for a range of sizes of the uncertainty standard deviation expressed as a percentage (%ESD) of the status class widths for the EQR (adapted from Clarke et al. 1996).

Confidence of class and implications of using multi-metric and worst case rules

Ralph Clarke finished the introductory section by pointing out that having sampling variation and other uncertainty in individual metrics and EQRs has consequences for status class assignments based on multi-metrics (Figure 6). Adding metrics which have very high sampling variances can make a multi-metric assessment less precise. Also, in applying the 'worst case rule' as recommended in the WFD for combining the WB assessments made using sample information from different BQEs, sampling variability and statistical implications of taking the worst (i.e. minimum) of our estimated individual metric classes for a WB, may cause a WB status class to be under-estimated (Figure 6).

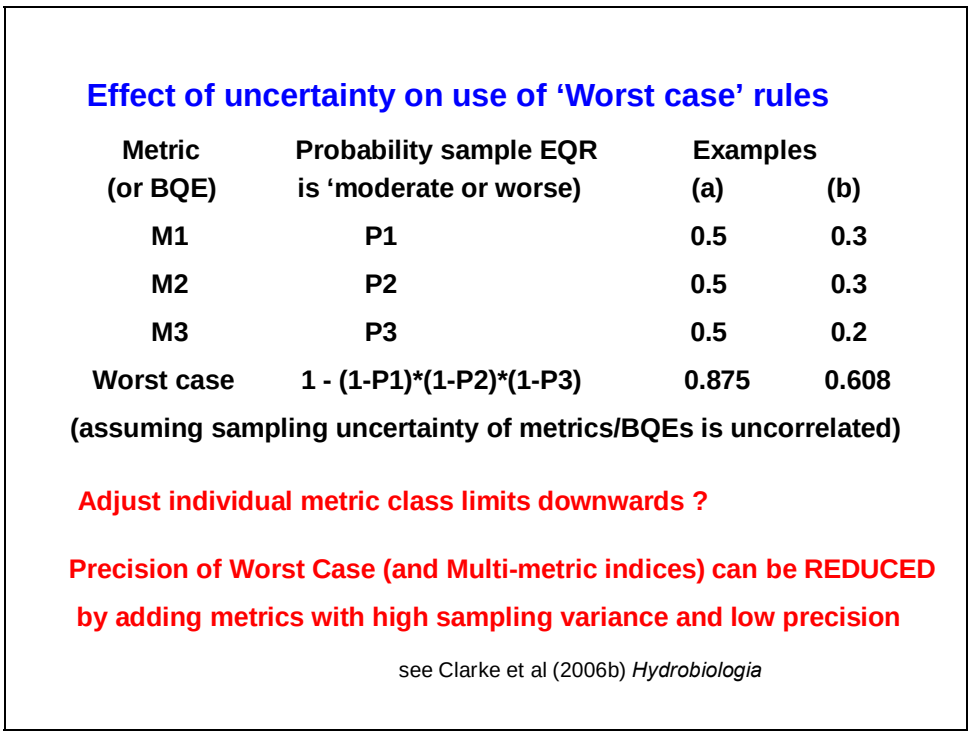

Figure 6 Consequences of uncertainty on multi-metric precision or applying the 'worst-case rule'.

Presentation by Ralph Clarke on sources of uncertainty and experiences from rivers

Note: The word "uncertainty" is mentioned in almost every WP description of the project's Description of Work and at a total of 163 times.

Biota-pressure-environment modelling and uncertainty

There are a wide range of potential approaches to making WFD-compliant WB biological assessments of WB ecological status and change in EQRs and status, but all approaches should involve a quantitative understanding of the major sources of uncertainty. The use of surveying and monitoring through sampling should obviously include studies and estimates the main factors responsible for variation between samples in both space and time. But it is also important to understanding the potential causes of lack of fit when using more mechanistic models of biota-pressure relationships and derived predictive estimates of the extent of pressure required to improve ecological status to 'good' or better. Ralph Clarke highlighted this by showing that if the statistically or mathematically modelled relationship between the observed sample biota and the pressure variable(s) is strong, then the correlatory (but not necessarily mechanistic) modelled relationship between the biota and the pressure variables must be strong and the sampling variability of the biotic dependent variable(s) must be relatively low (Figure 7). However, if the modelled relationship between the observed sample biota measure and the pressure variable(s) is weak, the cause could be that (a) there is only a weak underlying true relationship, (b) the sampling variability of the biotic measure is high or (c) both (Figure 7).

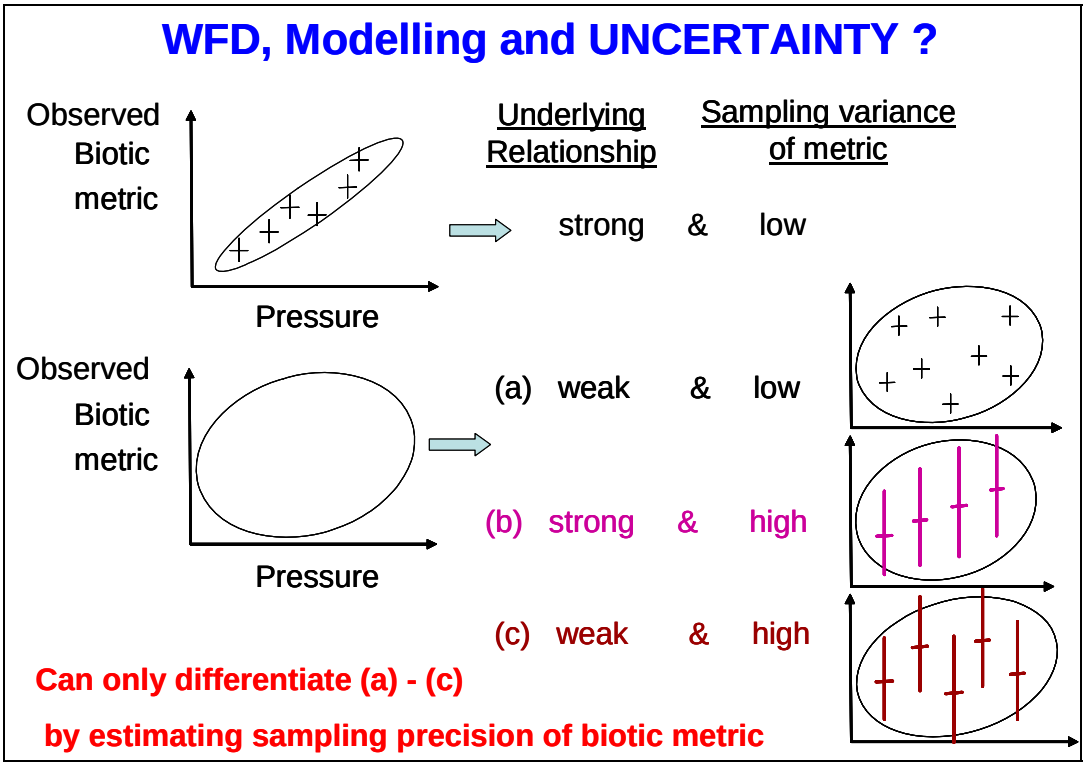

Figure 7 Sampling uncertainty influences strength of relationship between observed biota and pressures.

Summary of potential factors contributing to WFD status class uncertainty

Ralph Clarke then summarised the many potential factors which can contribute to the overall uncertainty and lack of confidence and reliability of ecological status class assessments and monitoring of change:

Sources of uncertainty in Observed biota & metric values

- Natural variation - Spatial (within sites and across WB)
- - Temporal (over time period to be reported)
- Sampling variation (method and personnel)
- Sample processing errors (sub-sampling, sorting, counting & identification)
- Effects of pollution or "environmental stress" or remediation to be detected

The aim and problem is to differentiate stress or remediation effects from 'natural' temporal variability (this is part of WISER Module 5)

Sources of uncertainty in setting Expected or Reference Condition (RC) biota & metric values

- Inadequate information & knowledge
  - Inadequate set of RC sites for all or some WB types
  - Not involving all "relevant" environmental variables
    (e.g. WFD System A or B Types or predictive model variables)
  - Not making optimum predictive model
    (e.g. RIVPACS v Neural Networks; mechanistic model functions/parameters)
- Sampling variation in RC sites' sample data (SE of mean)
- Inconsistent data – this is a potential problem in WISER
  - Existing Data from different sampling methods/standards combined to set RC
  - Test site's observed sample value and RC data values based on (partially) different sampling methods

Additional Sources of WB assessment uncertainty

- Choice of metrics – some biological metrics may have high statistical sampling precision, but not be very informative measures of WB ecological condition
- Choice of EQR measures
- Method to compare Observed (O) with RC biota and metrics
  (e.g. EQR = O / RC (as used in RIVPACS)
  or EQR = Likelihood of O given RC (as used in FAME EFI)
- Choices of status class limits for each EQR
  e.g. Setting high/good boundary ? use lower 10% EQR value of Reference sites
  or 'expert opinion' or maximise agreement with prior classes

uncertainty with estimation and use of WFD GIG inter-calibration factors

- Method for combining metrics, EQRs and status classes
  - use of Multi-metrics (e.g. river macro invertebrates ICMi = weighted mean)
  - worst-class or average/median class

Ralph Clarke made the point that it is useful to be aware that every decision you make can affect the estimates of status classes of WB and their uncertainty. However, he proposed that the practical way to progress is to

- Acknowledge that there is no absolute true or correct WB classification protocol
- Assess uncertainty due to natural and sampling variation conditional on the overall classification protocol choice of sampling method, metrics, EQRs and class definitions
- Try to compare different classification protocols (methods, metrics, and BQEs) and learn from their discrepancies and improve

Experience in assessing uncertainty in river bioassessments based on RIVPACS

The RIVPACS (River Invertebrate Prediction And Classification System) approach to assessing the biological of river sites is based comparing the observed biota (O) with the biota expected (E) under high quality or unstressed reference conditions at sites of that physical type, as predicted from a statistical model relation the environmental characteristics of a set of carefully selected high quality reference sites to their observed biota sampled in a standardised way. RIVPACS was slowly developed and improved over many years, mostly prior to the introduction of the WFD in 2000 - which it may have influenced (Moss et al. 1987, Furse et al. 1995, Moss et al. 1999, Wright et al. 2000; Clarke et al. 2002, Clarke et al. 2003).

Ralph Clarke described how, in the early 1990s, following the first national river survey based on RIVPACS, the UK RIVPACS team were asked "What are the confidence limits for the O/E ratios, what confidence can we have in the derived estimates of biological quality class, and how can we tell whether sample estimates of O/E and quality class taken in different years from a river site indicate real change in site quality?" This type of question about uncertainty is still at the heart of many of our research and assessments method developments in the WISER project, whether for rivers, lakes or coastal and transitional waters and whether based on macroinvertebrates or other BQEs.

Replicated sampling study across a range of site types and qualities

Ralph Clarke described how to start assessing uncertainty for the RIVPACS sampling method and system, the then Institute of Freshwater Ecology (IFE) RIVPACS development team conducted a carefully designed and balanced replicated field sampling study involving:

4 contrasting RIVPACS (TWINSPAN) stream types
x 4 quality classes (A, B, C, D) = 16 sites
x 3 seasons (spring, summer, autumn)
x 3 replicate samples (2 x IFE person, 1 x local Agency biologist)
(samples 1 and 3) (sample 2)

Sampling variability in biological metrics is likely to depend on the type or quality of site, so it was, and is, considered important to assess sampling variability across a range of physical types and qualities of site, not just amongst high quality reference sites. The variance between replicate sample values of the metric 'Number of BMWP taxa' (TAXA) was found to increase with the mean of the replicate values for a site. By working with the square root transformed values of observed TAXA (√TAXA), the variance between replicate sample values became roughly constant (Figure 8).

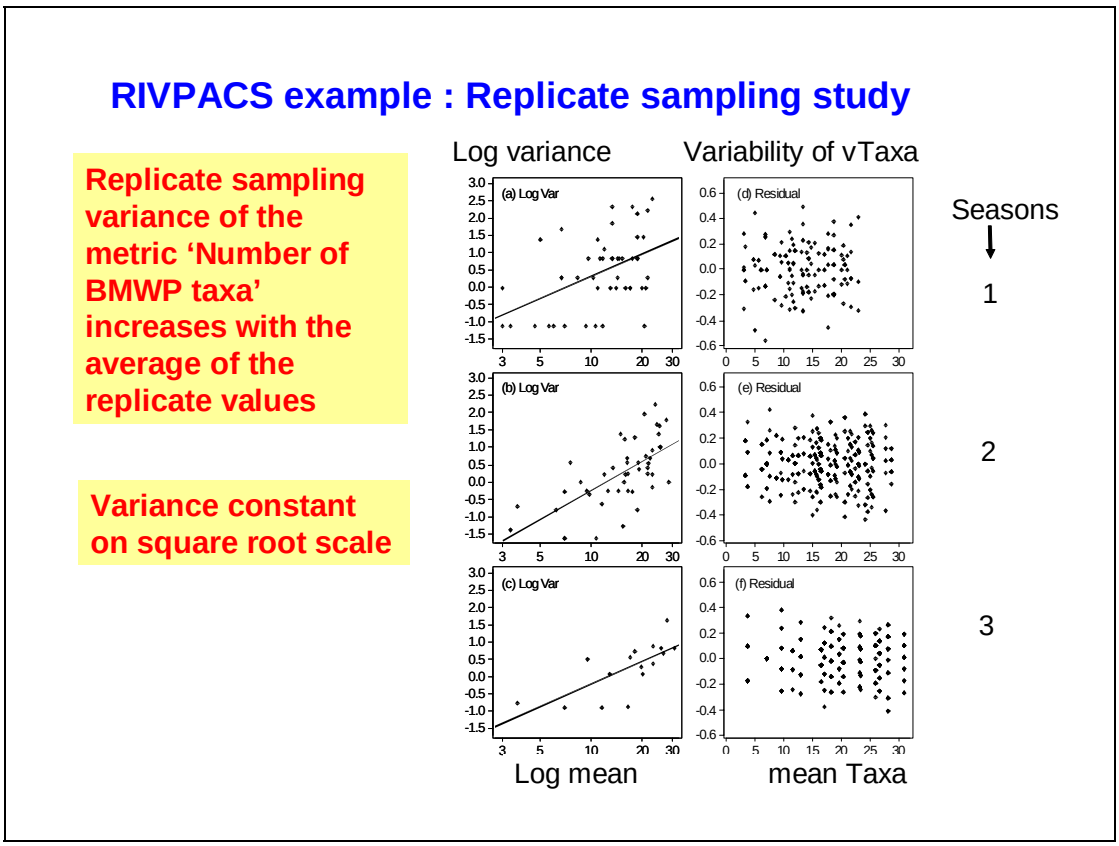

Figure 8 RIVPACS replicate sampling variance increase with mean but is constant on a square root transformed scale.

The replicate sampling variance for the metric ASPT (BMWP 'Average Score Per Taxon') was found to be independent of site type or quality. It was also independent of the number of scoring taxa present, but Ralph Clarke warned that this is not necessarily true for other taxon weighting metrics and the pattern and size of sampling variability in each individual metric to be used in WB bioassessments needs to be assessed. The sampling variability (SD) of √TAXA and of ASPT only depended on whether the values were calculated for single season samples, or two or three season combined samples as often used by the UK government environment agencies in national surveys and long-term monitoring of UK rivers using RIVPACS (Table 1) – see Clarke et al. (2002) for further details.. In the mid 1990s, this and related studies helped the UK agencies decide that it was most cost-effective to base national surveys of river quality on two season (spring and autumn) combined sampling, eliminating the cost of the previous additional summer sampling campaign.

Table 1 Estimates of RIVPACS replicate sampling standard deviations (SD) for the metrics (a) square root of the number of BMWP taxa (√TAXA) and (b) ASPT for (1) single season samples (2) two season and (3) three season combined samples

	Number of seasons samples combined for site assessment		
Metric	1	2	3
√TAXA	0.228	0.164	0.145
ASPT	0.249	0.161	0.139

Assessing inter-operator effects and reliability of change with staff turnover

Ralph Clarke said that the extent of sampling operator effects can be assessed by comparing different in metric values between replicate samples taken by the same person with differences in values between replicate samples taken by different people at the same site on the same day. In the RIVPACS study, inter-personnel differences were minor or negligible with statistical analysis of variance components showing that operator variance accounted for only 4% and 12% of the total (operator + true replicate) variance in ASPT and √TAXA respectively.

Ralph Clarke pointed out obtaining small differences between operators (i) may be dependent on having well trained personnel and (ii) is crucial for our confidence in wide-scale long-term monitoring of change where there is likely to be a turnover of staff, as occurs with the national and regional environment agencies. We need to minimise the risk that observed sample changes in biological metrics between years are real and not merely due to being sampled by different staff!

Sample processing sorting and identification errors

In sorting and identifying the taxa in a new sample some taxa may be missed or mis-identified by less experienced staff. This may lead to biases and under-estimation of the site's O/E ratio for number of taxa (TAXA) and maybe ASPT.

Ralph Clarke described how, since the early 1990s, IFE/CEH have been contracted by the UK government environment agencies to re-analyse a proportion of all UK RIVPACS samples to provide an external audit and quality check. Such biological audits can provide annual estimates of the distribution of sample processing errors, highlights taxa which are most frequently missed or mis-identified and provides estimates of the biases in O/E values due to missing taxa for each Agency sample processing laboratory or region. Ralph Clarke also derived relationships between the ASPT of the missed taxa and the recorded taxon richness which he then used in the RIVPACS III+ software system for site assessments to correct for the current level of sample processing errors within a lab. Without this correction, different between years in sample processing accuracy would be another source of apparent variation and changes in site quality over time. Ralph Clarke said that since auditing begin the number of processing errors has been reduced – this may be due to simply having an audit scheme in which people know they may be checked, but also by feed-back from the identification results from the scheme and related staff

training. Ralph Clarke reported that Peter Haase of Senckenberg Research Institute, Germany, is currently developing a similar macroinvertebrate sample audit scheme for river sites in Germany.

Errors in the Expected (E) fauna for RIVPACS

When considering errors in the RIVPACS expected values: the choice of reference sites, environmental predictor variables and statistical prediction method used in RIVPACS were all treated as an integral part of the definition of the condition index. The errors in the Expected fauna were assumed to arise only from errors in measuring the environmental predictor variables for new sites. Their effect was quantified using independent measurement by all personnel for each of our replicated study sites.

Computer-based sensitivity analyses were also used to assess the sensitivity of RIVPACS expected values for BMWP indices to errors or variation in each of the RIVPACS environmental predictor variables. Ralph Clarke reported that variation between people in their recording of stream width, depth and substratum in the field were all well within the acceptable tolerance limits determined by the sensitivity analysis.

Similar types of computer-based sensitivity analysis of predictive or mechanistic models could be useful for other bioassessment approaches, WB types and BQEs.

RIVPACS Uncertainty simulation model

Ralph Clarke showed how estimates all of these above sources of variation, errors and uncertainty are combined within an uncertainty simulation model within the RIVPACS software (Figure 9).

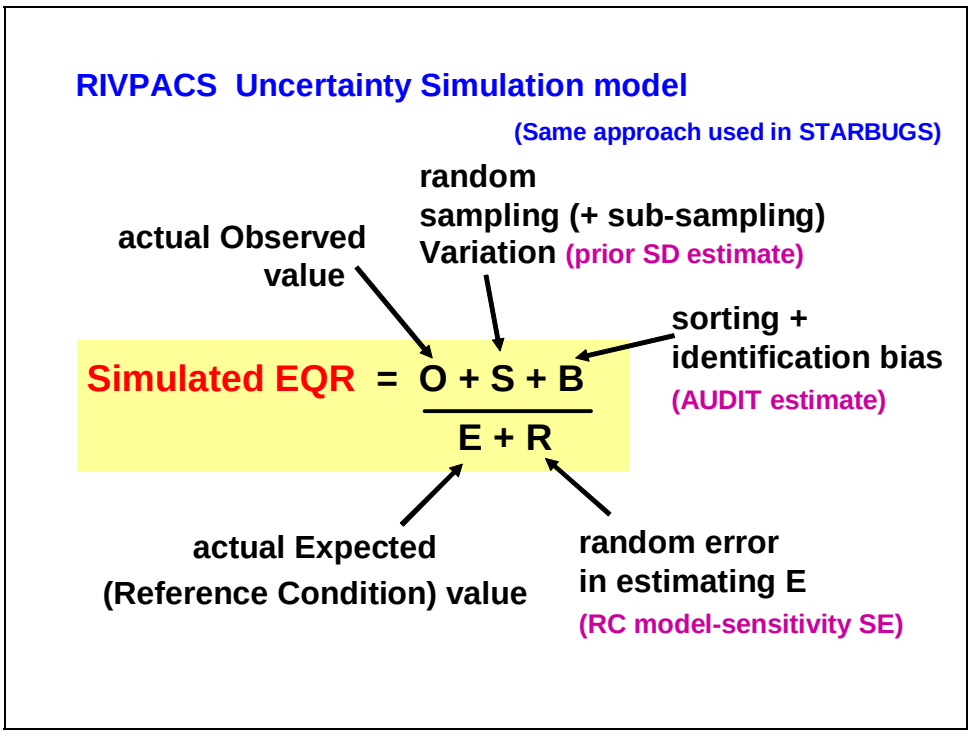

Figure 9 Uncertainty simulation model to quantify EQR uncertainty and confidence of class.

Such Monte Carlo simulation models any us to estimate confidence limits for sample EQR values, confidence of class, confidence of real change in average sample EQR for a site and confidence of a real change in status class based on either single metrics (as shown in Figure 10) or more complex multi-metric or worst case rules – further details are given in Clarke (2000).

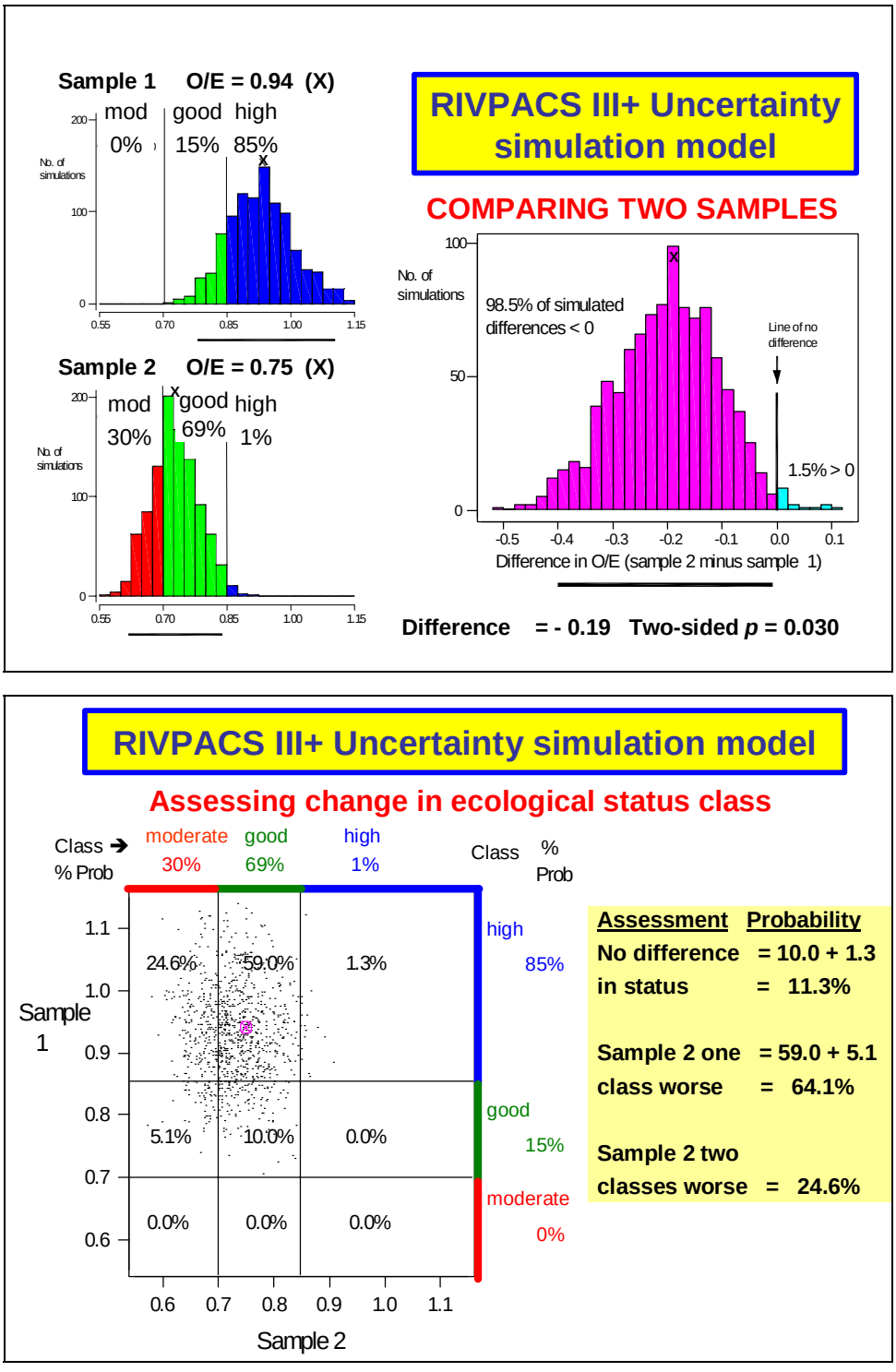

Figure 10 Assessing change between two samples using the RIVPACS Uncertainty simulation model.

Ralph Clarke said that he had initiated a similar type of uncertainty simulation approach within the STAR project as STARBUGS and this will be adapted for use within our WISER project as WISER Deliverable D6.1-3.

Implications of spatial and temporal scale of WB bioassessment

Ralph Clarke then used the example of very recent extensions to UK RIVPACS and its uncertainty estimation to highlight how the confidence of a sample estimate of ecological status class depends on the spatial and temporal scale over which the assessment estimate is to apply.

In the past, UK national survey monitoring of long-term biological quality of a river stretch using macroinvertebrates and RIVPACS was based on sampling a single site (in spring and autumn combined) every one in five years. The sampling site was assumed to be representative of the whole stretch; stretches being formed to be as homogeneous as possible. As the sample-based site assessment was effective just for that site in that year and RIVPACS predictions are both site- and season-specific, inter-season variation in observed fauna was considered to be controlled for and therefore the only factors which were allowed for in the assessment of uncertainty of status class for that site that year were (i) replicate sampling variability and (ii) sample processing errors/biases (sorting + identification), based on prior estimates derived from the replicated study and the audit.

However, it is proposed that future UK national monitoring for the WFD will be for WFD Water Body units which will generally be for longer river sections for management (i.e. larger spatial scale) and be based on estimates of average quality across the WB over a 3-year reporting period. Uncertainty then depends on:

Replicate sampling variance and sample processing errors as before

Temporal variance (both short-term within-season and between-year (within 3-year period) irrespective of whether it is “natural” or anthropogenic due to changes in stresses and their impacts

Spatial variance between potential sampling sites with the WB

+ Spatial – temporal interaction (i.e. site x year), measuring the extent to which changes in quality between years vary between sites within the WB)

Ralph Clarke said that as there was no single historical dataset on UK rivers from which all of the various variance components could be estimated, he had to combine datasets from different regions and studies. This makes statistical estimation of the variance components by analyses of variance, mixed model and maximum likelihood techniques both more difficult and less reliable. This is because the estimates of all higher level components (e.g. inter-year variance) are dependent to some extent on the recorded variability and estimates of variance of lower level components (e.g. replicate and short-term temporal variance). In general, it should be remembered that estimates of variance components can themselves be imprecise and dependent on the sites and years data available. Ralph Clarke then explained how the overall variance of an

estimate of the mean value of a metric or EQR for a WB over a time period depends on all of the various sources of sampling variability and the extent to which our sampling scheme encompasses and controls for the various aspects (Figure 11).

Uncertainty variance depends on sampling scheme

Example : Observed sampling scheme mean ASPT value for WB = 6.4

Good/Moderate boundary for this WB type = 6.0

Observed class is 'Good'; Probability of Failure (Moderate or worse) ?

Estimates from previous study

ASPT	Reps	Dates/year	Years/period	Sites/WB	Total	Probability
Variance	0.065	0.059	0.021	0.105	0.250	Failure
Scheme 1	1	1	1	1	0.243	21%
Scheme 2	3	1	1	1	0.200	19%
Scheme 3	1	1	3	1	0.146	15%
Scheme 4	1	1	1 different site each year		0.076	7%

$$\frac{Variance}{of\ .mean} = \frac{V_{Rep}}{rwsy} + \frac{V_{WSeas}}{wy} + \frac{V_{Year}(1-y/3)}{y} + \frac{V_{Site}}{s}$$

where r = replicate samples per site per date

w = dates sampled per year (within specified season(s))

y = years sampled (1, 2 or 3) in 3-year period

s = sites sampled per water body

	v Observed TAXA		Observed ASPT	
Variance component	**Var**	**% total within WB**	**Var**	**% total within WB**
Replicate sampling	**0.058**	**37%**	**0.065**	**26%**
Season	**0.035**	**23%**	**0.059**	**24%**
Between Years within period	**0.036**	**24%**	**0.021**	**8%**
Between sites within WB	**0.025**	**16%**	**0.105**	**42%**

Figure 11 Estimation of spatial, temporal and replicate variance components and implications for risk of mis-classification using a range of sampling schemes.

Experiences from the EU FP5 STAR project

In his final section Ralph Clarke highlighted the some of the experiences of assessing sampling method variability within the European FP5 project STAR (STAandardsiation of River classifciations) led by Mike Furse of CEH. Ralph Clarke said as that national macroinvertebrate sampling methods had already been established prior to STAR, the project team took the pragmatic view that it was likely that individual member states would continue with their existing methods for long-term monitoring compatibility. Therefore in order to assess the relative susceptibility of each national sampling method and sample processing protocol to uncertainty, at each study site, project partners took samples using both the national method and a standard STAR-AQEM method based on the AQEM method developed during the previous AQEM project (led by WISER project leader Daniel Hering) (Furse et al. 2006). Where an agreed 'national' method was unavailable, the partner collected and processed samples using the RIVPACS sampling protocol. The study sites in the STAR field sampling campaign were carefully chosen to cover a range of stream types and a range of perceived biological qualities within each type. Surprisingly the AQEM project had not involved any assessment of sampling variability.

Ralph Clarke explained that, in order to assess the susceptibility of each method to sampling variability and to compare methods, STAR partners took two replicate samples in each of the two sampling seasons at a subset of 2-6 study sites within a total of 18 stream types spread over 12 countries using both the STAR-AQEM protocol and either the 'national' or RIVPACS protocol. All macroinvertebrate metric sample values were calculated using the AQEM/STAR ASTERICS software to ensure compatibility.

Ralph Clarke said that comparing the relative replicate sampling precision of different methods and of different metrics with widely different forms and units from richness to percent composition of certain taxonomic groups can be tricky. His solution was to use analysis of variance techniques to estimate the variance components (replicate, between sites, between seasons, between stream types) and then calculate the replicate variance in a metric as a percentage (P_{samp}) of the total variance in the metric's values within a stream type and season (Figure 12). Low values of P_{samp} indicate that for that method and metric, replicate sampling variance is only a small fraction of the total variability across streams of differing perceived qualities in the same stream type and the sampling precision of the metric is high with the potential to also have statistical power to detect the effects of environmental stress on stream quality. In contrast, method and/or metrics with very high values of P_{samp} cannot have high power to detect effects of stress replicate sample values within a site are almost as variable as values from different perceived quality sites.

Obviously high sampling precision is a requirement for a metric to be effective, but it does not necessarily indicate that the metric has statistical power to discriminate between stress levels and indicate stream quality status. Ralph Clarke thought that the use of measures like P_{samp} could be useful in our WISER project for comparing metrics and/or sampling/surveying methods from different BQES for lakes, coastal or transitional waters.

Compare % Sampling precision of metrics

Metric	f(x)	SD	P_{SAMP}
Abundance [ind/m²]	vvx	0.596	19
Number of taxa	vx	0.364	21
Number of Families	vx	0.283	20
Number of EPT taxa	vx	0.273	8
Saprobic Index	x	0.055	6
German Saprobic new index	x	0.050	4
ASPT	x	0.227	14
Diversity (Shannon-Weiner)	x	0.237	18
% Rheophilic	asin	0.075	7
% Littoral	asin	0.031	3
% Grazers/Scrapers	asin	0.041	8
% Shredders	asin	0.059	17
% Gatherers/Collectors	asin	0.054	14
% Oligochaeta	asin	0.075	25
% EPT individuals	asin	0.070	10
% EPT Taxa	asin	0.042	14
Trait m1 : max body size = 1cm		0.023	14
Trait m7 : crawler locomotion		0.029	26
Trait m12 : current <25cm/s		0.014	4

P_{SAMP} = sampling variance as % of total variance in metric values within a stream type

SD = Mean replicate sampling SD

Based on RIVPACS sampling in Austria, Germany and UK

f(x) = optimum transformation to make sampling SD homogeneous

See Clarke & Hering (2006)

Clarke *et al* (2006 a,b)

STAR issue of *Hydrobiologia*

Comparing Sampling precision of methods

Sampling variance as % of total variance in metric values within a stream type (overall average)

P_{SAMP}

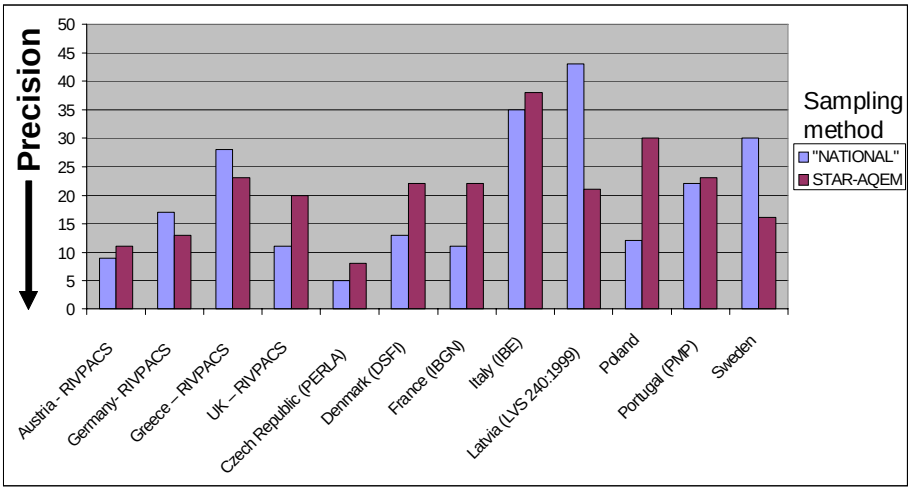

Figure 12 STAR project comparison of relative sampling precision of macroinvertebrate metrics and methods.

Ralph Clarke recommended that WISER partners refer to the special issue of the journal *Hydrobiologia* devoted to the results of the STAR project (Furse et al. 2006), and especially to the section of papers on 'Errors and uncertainty in Bioassessment methods' introduced and summarised by Clarke and Hering (2006).

Presentation by Iwan Jones on uncertainty assessments based on experience with deep waters and lakes

The ideal methodology, sampling strategy and metric will produce an assessment perfectly correlated with true differences in ecological quality, despite variation in the sample composition due to a variety of spatial and temporal factors. Such a tool does not exist: we have to be able to determine how much of the variation in assessments is due to true variation in quality and how much due to other factors. Uncertainty must be quantified and wherever possible, minimised. Only in this way can we know how confident we are that an assessment of a WB reflects its true ecological quality class.

Iwan Jones presented an idealised design for the quantification of uncertainty due to sample collection and processing. If we understand which sources of variation make substantial contributions to the uncertainty associated with an assessment, monitoring strategies can be designed to reduce this uncertainty and give an acceptable level of confidence.

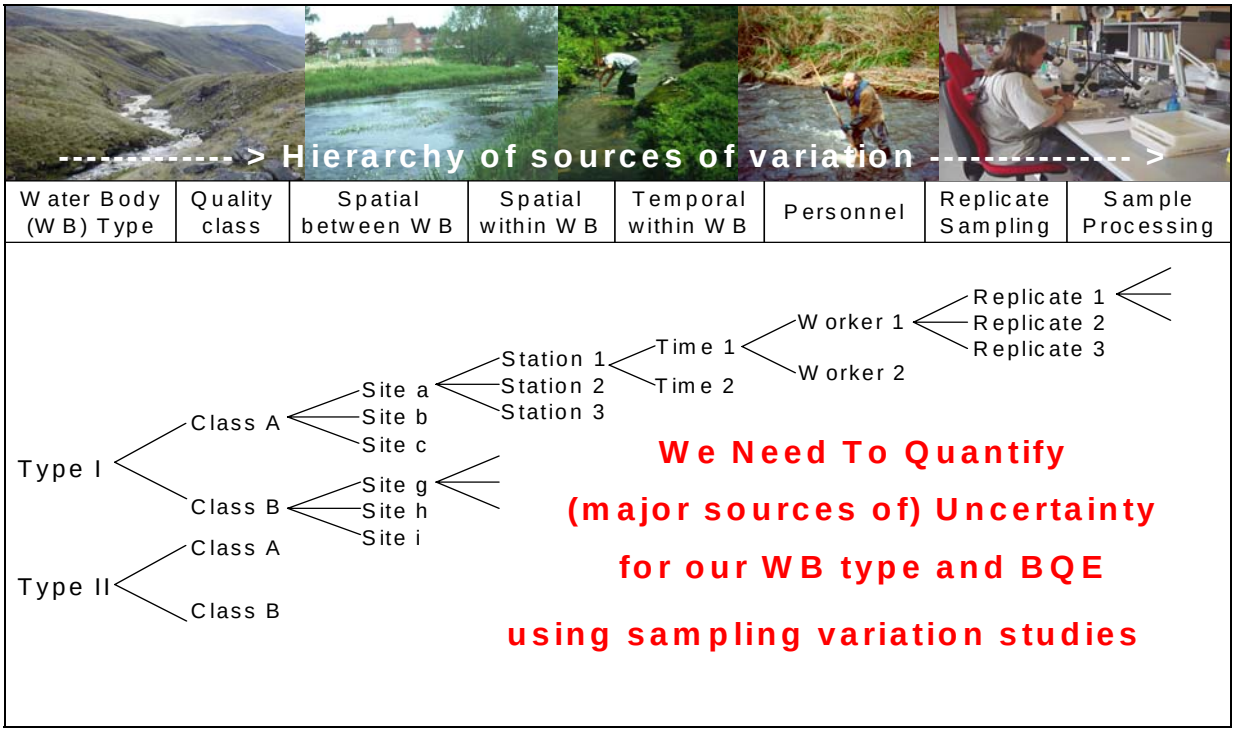

Figure 13 An idealised design for the quantification of uncertainty due to sample collection and processing.

Cost-effectiveness of different sampling techniques

Uncertainty is affected by factors other than spatial and temporal variation, such as the choice of methodology to assess the BQE and the choice of metric to use. A full understanding of uncertainty can help guide selection of methods to be used to make assessments of quality in the most cost-effective manner. Iwan Jones illustrated this with a presentation on uncertainty and sampling method selection in deep rivers which can be difficult, dangerous or impossible to

sample by the common pond-net and kick sampling methods. At each site, different sampling methods (air-lift, dredge, margin hand net, long-handled pond net) were used by different workers, with three replicate samples collected (two by one worker one by another). The effects of different sources of uncertainty were illustrated as within site, between site, worker-dependent and replicate effects on the BMWP/TAXA/ASPT indices calculated for the samples. Together with cost estimations for each method, uncertainty was used to evaluate the cost-effectiveness of each method (Table 2). Using the average cost to make a sample assessment by different techniques (here time to process samples) and the number of sample assessments required to achieve a set level of confidence in the result (i.e. percentage SE of the mean metric value), it is possible to compare the cost-effectiveness of the different techniques. This approach is liable to be useful when comparing novel "quick and dirty" techniques (such as use of aerial photography) to traditional techniques. Ideally all costs should be taken into account (i.e. equipment costs (e.g. boat hire), time to collect a sample, time to process samples and identification skill/costs of required staff).

Table 2 Cost (total time taken) to achieve set level (10%) of variance attributable to sampling when sampling deep river macroinvertebrates using four different techniques (airlift, light dredge, marginal sweep and Long handled pond net). (ASPT = BMWP Average Score Per Taxon)

	Technique			
	Airlift	**Dredge**	**Margin**	**LHPN**
Cost per sample (min)	*267*	*94*	*147*	*102*
(a) BMWP	801	∞	1029	510
(b) TAXA	801	2914	1617	612
(c) ASPT	801	15040	441	1428
In all 3 metrics	**801**	∞	**1617**	**1428**

Assessing sources of variability in littoral lake invertebrates

Iwan Jones went on to describe how an understanding of uncertainty can be used to guide the design of sampling strategies, using lake macroinvertebrates as an example. In this case a more confident whole lake assessment will be made by distributing samples across more stations rather than taking more replicates at each station (Figure 14). He then described how sample processing errors can add uncertainty (typically bias) to assessments. This applies to all BQEs; the realisation that a BQE is present, that its taxonomic identity may be different to those individuals already identified, and correctly ascribing an identity are prone to error and hence have the potential to influence uncertainty. Training can be used to reduce uncertainty, but variation among assessments made by different workers, both sample collection and processing, is part of the uncertainty in any assessment. Quality assurance can be used to determine bias due to processing mistakes. It should be remembered that the EQR is dependent upon the probability

of occurrence of individual taxa in both reference and observed data: the quality of the reference condition data is paramount.

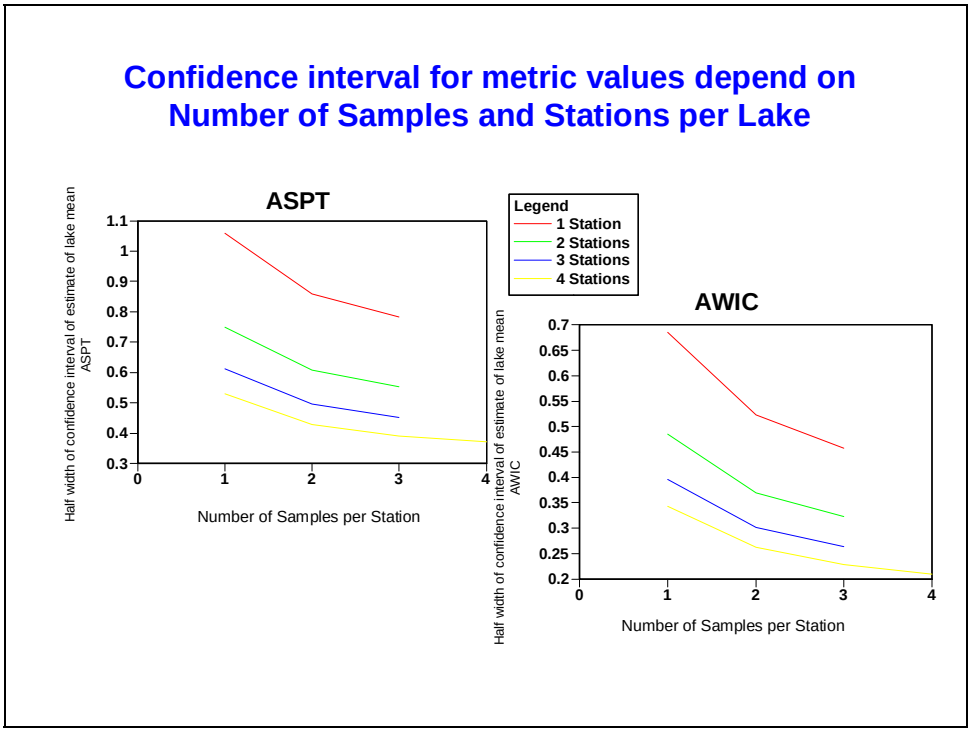

Figure 14 Relationship between confidence in whole lake assessment, number of stations sampled around lake perimeter, and number of samples collected per station using lake macroinvertebrates and two indices BMWP Average Score Per Taxon (ASPT) and Acid Waters Indicator Community (AWIC).

Phytoplankton uncertainty and ring tests

Finally Iwan Jones described a recent study with Laurence Carvalho (CEH) on sources of phytoplankton community and metric uncertainty. Within lake variation incorporates variation due to:

- Sampling locations, both horizontally and vertically (spatial variability)
- Sampling variability
- Sub-samples and fields of view observed (sub-sample variability)
- Years, seasons, months, days, hours (temporal variability)
- Observer or analytical error ("counter" variability)

The study focused on importance of sub-sample/counter variability and the question "Can uncertainty be reduced by training?"

If in a ring test, one sample was taken per lake and passed between workers who each processed a different sub-sample, you can't tell whether differences between workers are due to sub-sampling variability or "counter" variability.

In the recent study, replicate sub-samples were taken from different stations within each of a range of lakes and different 'counters' processed the same sub-samples and variance components estimated and compared for importance (Table 3). In this example differences between counters accounted for about 10% of the total within lake variability in a phytoplankton metric. Different metrics quantify different aspects the BQE being assessed and, hence, do not necessarily respond to variation in the same way. The choice of metric to be used will influence the uncertainty associated with the assessment, and hence the confidence of class (Table 3).

Table 3 Variation in lake phytoplankton metrics attributable to different sources in sample processing. Metrics do not perform equally. The percentage of the variation attributable to counters may be reduced with training; $P_C = 100\ V_C / V_W$, $P_W = 100\ V_W / V_S + V_C + V_B$..

Metric	Sub-sampling V_S	Counter bias V_C	Within-lake V_W	Between-lakes V_B	**% counter** $\boldsymbol{P_C}$	**% within** $\boldsymbol{P_W}$	**Sample processing SD** $\boldsymbol{SD_W}$
Observed Score	0.00064	0.00007	0.00071	0.00274	**10**	**21**	**0.027**
EQR Score	0.01199	0.00154	0.01353	0.02155	**11**	**39**	**0.116**
$Log_{10}Log_{10}$Total Biovolume	0.00146	0.00002	0.00148	0.00367	**1**	**29**	**0.038**
Log_{10} Total Taxa recorded	0.00653	0.00096	0.00749	0.02296	**13**	**25**	**0.087**
Log_{10} Taxa Matched	0.0064	0.00284	0.00924	0.02664	**31**	**26**	**0.096**

Presentation by Jacob Carstensen on uncertainty quantification from monitoring marine WB

Jacob Carstensen showed examples of uncertainty quantification from marine monitoring data. He presented a schematic view of how factors causing uncertainty propagate themselves up through all of the stages of water body status class assessment and up management biota-pressure management models (Figure 15). Jacob Carstensen pointed out that the CIS Guideline13 only requires sampling and measurement error to be included in assessments of confidence in estimates of WB status class

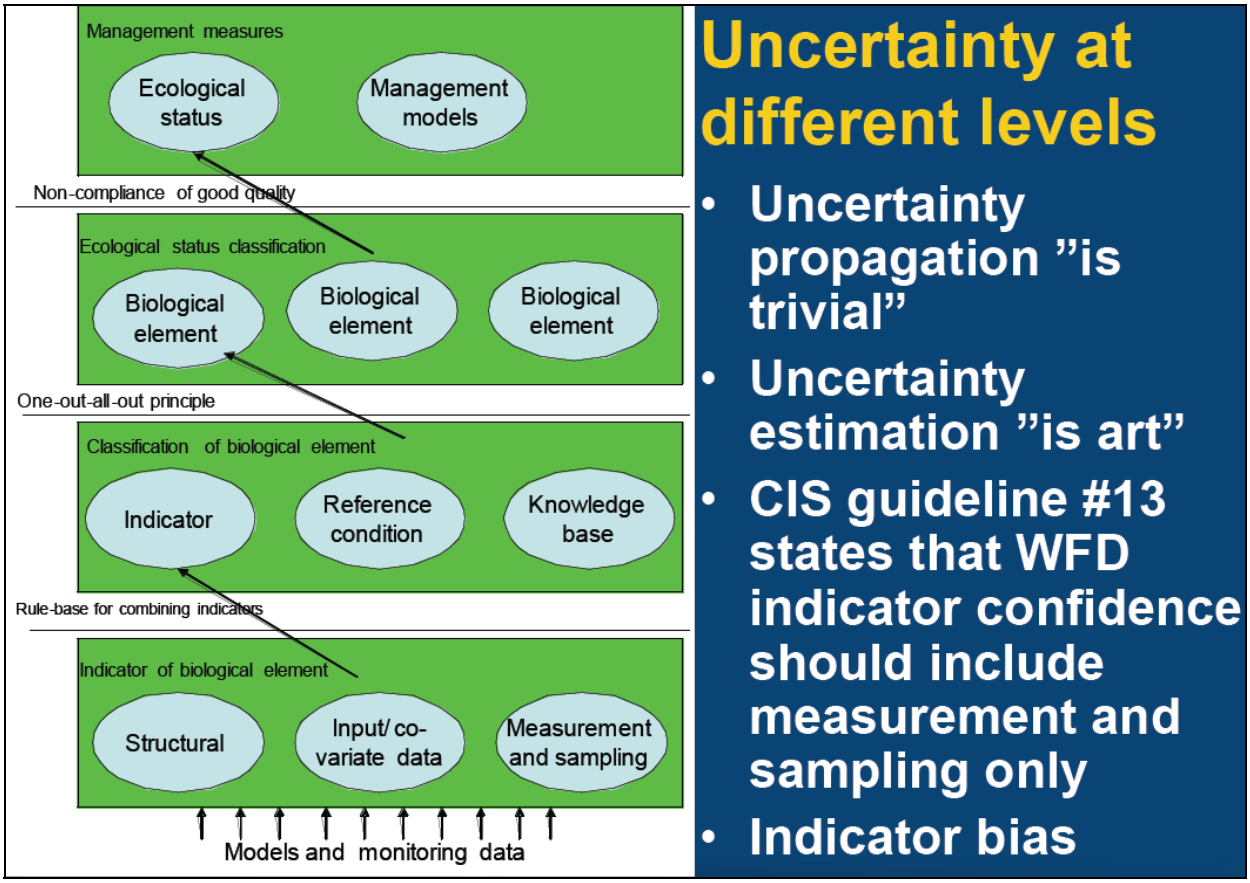

Figure 15 Schematic view of propagation of uncertainty in water body bioassessment and management.

Jacob Carstensen made reference to his paper (Carstensen 2007) on 'statistical principles for ecological status classification of Water Framework Directive monitoring data', published in a special issue of Marine Pollution Bulletin edited by Devlin Best and Haynes (2007), which is highly relevant to our WISER project and to which WISER partners are referred for further details.

Jacob Carstensen pointed out that statistical power analyses of sample marine data often indicates that a prohibitively high number of samples are required to give sufficiently precise

estimates of biotic and nutrient metrics considered to be potential useful indicators of marine WB ecological status (Figure 16).

Number of hydrochemistry observations required (α=5%, β=20%)

Variable	Deviation from boundary				
	d'=5%	*d'*=10%	*d'*=15%	*d'*=20%	*d'*=25%
Ammonium	2123	557	259	152	102
Nitrate + nitrite	1454	381	178	105	70
Total nitrogen	124	33	16	9	6
Phosphate	1156	303	141	83	56
Total phosphorus	306	81	38	22	15
Chlorophyll *a*	1291	339	158	93	62
Carbon biomass	3411	894	416	245	164

Figure 16 Number of sample observations required for correct classification with a power of 80% if the true mean deviates by d_0 from a classification boundary for the seven variables (α = 0.05) and standard error (s) of the mean is estimated from all stations combined)- from Carstensen (2007).

Jacob Carstensen gave an illustrative example of the complex spatial-temporal biological variability which can occur within a individual marine water bodies, which explains why such high sampling intensities can be needed to give adequate precision of say annual mean chlorophyll concentration for the whole WB (Figure 17).

Jacob Carstensen suggested that one cost-effective way to improve precision through the use of statistical modelling techniques to try to understand and control for variation due to systematic seasonal variation and other influential covariates factors (Figure 18).

The take-home messages were:

- There should be increased use of modelling and new sampling techniques for reduce random variation
- All "available data" should be used for indicators
- Metric and uncertainty parameters should be estimated from a mixture of models of monitoring data and experiments designed to estimate uncertainties

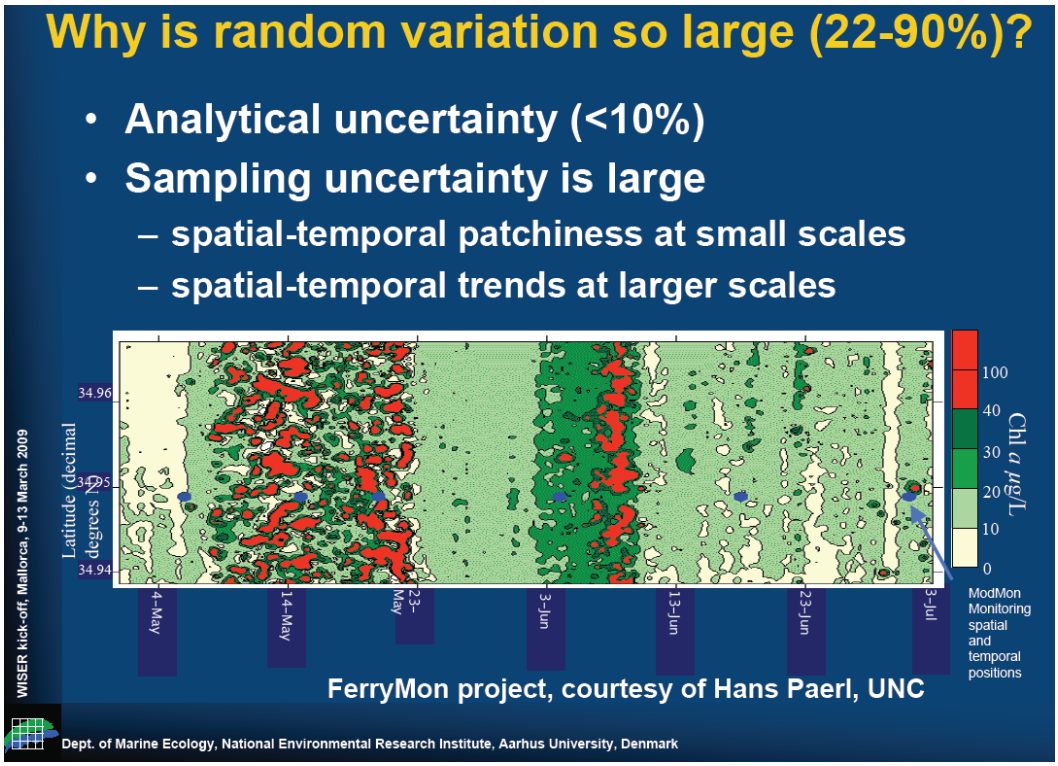

Figure 17 Example of high small scale spatial-temporal patchiness and larger scale trends in chlorphyll-a concentrations within a marine WB.

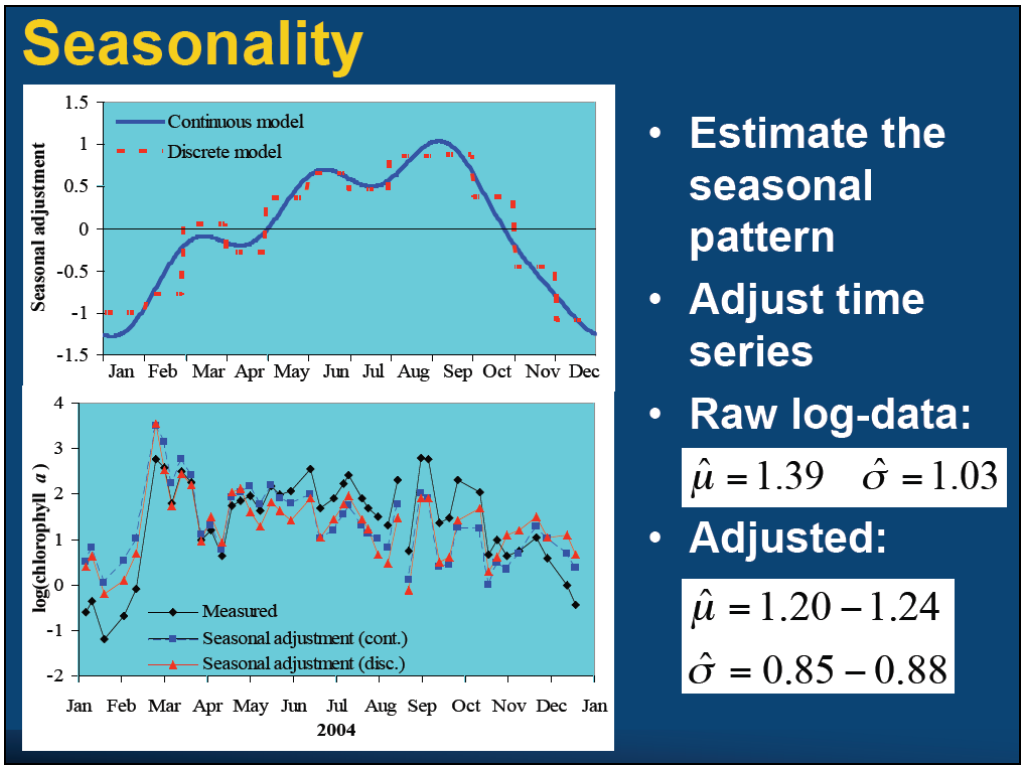

Figure 18 Example of seasonal adjustment modelling of log chlorophyll-a concentrations from a monitoring station at Limfjorden estuarine complex in Denmark.

Breakout groups to consider how to assess uncertainty within each WP

Following the formal Uncertainty workshop presentations, each WP within Modules 3–5 formed a small breakout group to identify the major sources of variation and uncertainty and to discuss the most effective and affordable sampling programme and analytical approaches to measure and quantify some of these aspects of uncertainty.

Provisional plans of each Module and WP for the field sampling programme involving assessment of sources of sampling variation

As a result of discussions during the uncertainty workshop and throughout the WISER Start-up meeting, plans for the field sampling campaign were improved to provide better and more comprehensive study of some of the perceived major sources of uncertainty. Most WP now included field sampling plans to assess spatial variability within water bodies together with variability between WB for each of several combinations of pre-assessed quality class and types of lake and coastal/transitional water. Figure 19 gives the state of draft field sampling plans shortly after the workshop and WISER Start-up meeting for agreed modification by the individual WP leaders.

Updated details of the WISER field sampling plans and progress are posted on the WISER Intranet for partner information and collaboration.

Processes and responsibilities for assessing uncertainty agreed at WISER Start-up meeting

General

- Sampling and uncertainty data collected and stored by each WP
- Data coding must identify individual components of the sampling hierarchy (site, sample, method, replicate, worker…)
- If cost-effectiveness is to be assessed, cost (processing time) must be recorded
- Samples collected external to WISER must be compatible methodologically and must follow the same hierarchical sampling and data coding scheme
- Analysis of sampling variance will be performed by each WP individually
- WP6.1 can provide advice on sampling variance estimation, but only within its very limited resources

WP6.1 Responsibilities

- Review the current status of estimation of uncertainty
- Develop guidelines on the inclusion of uncertainty in the development and validation of indicators and develop methods to estimate the uncertainty of the different techniques and BQEs
- Develop guidelines for quality assurance, harmonisation of methodologies and taxonomy and the assessment of procedural bias in sample processing
- Advise on the design of data collection and statistical analyses for the uncertainty of metric data obtained in Modules 3 and 4
- Produce a generally applicable uncertainty estimation tool (based on the STARBUGS software tool developed in STAR) for assessing confidence of status classes

The first four of these bulleted responsibilities have already be partially achieved through the WISER uncertainty workshop and this report.

Change of WP leader

In May 2009, Iwan Jones left CEH and joined Queen Mary University of London (QMUL) and is therefore no longer formally involved in WISER. His CEH role as formal Leader of WISER WP6.1 has been taken over by Mike Dunbar of CEH, whom some will no from his involvement in the REBECCA and other projects.

WP6.1 formal Deliverables

D 6.1-1: Report on a workshop to bring together experts experienced with tool development and uncertainty estimation (Month 6, Lead person: Ralph Clarke (BourneU) – this document

D 6.1-2: Manuscript reviewing components of uncertainty and their assessment, including guidelines for estimation and quality assurance (Month 30, lead person: CEH)

D 6.1-3: Generally applicable software tool for assessing confidence of status class (Month 34, Lead person: Ralph Clarke (BourneU))

Water category	Work package	Water body type	Quality class	Spatial between water bodies (sites)	Spatial within water bodies (station/transect)	Temporal within water bodies (season/week/day)	Operative (worker)	Sampling and processing (replicates)
Lakes	Phytoplankton (WP3.1)	3 water bodies (lakes)	3	3 water bodies (lakes)	3 stations, 2 replicates per station	using existing data (not part of field campaign)	2	4 sub-samples per station. 3 from 1 replicate sample, 1 from other sample
	Macrophytes (WP3.2)	3 lake types	3 quality classes (only eutrophication)	3 water bodies (lakes)	6 localities	-	-	3 replicate transects
	Invertebrates (WP3.3)	3 water bodies (lakes)	3 quality classes related to morphology	3 water bodies (lakes)	Reference lakes: 3 stations each at wind-exposed and wind-sheltered shores, Morphologically altered lakes: 3 stations at 3 shoreline types	using existing data (not part of field campaign)	-	using existing data (not part of field campaign)
	Fish (WP3.4)	3 water bodies (lakes)	to be defined	3 water bodies (lakes)	using existing data (not part of field campaign)	-	-	using existing data (not part of field campaign)
Transitional and coastal	Phytoplankton (WP4.1)	5 coastal, 3 transitional	-	-	4-12 stations per type	using existing data (not part of field campaign)	parallel analysis of 2 workers	3 replicates per station
	Macroflora (WP4.2)	Macroalgae: 2 coastal types, 2 trans. types; Angiosperms: 1 coastal, 2 trans.	using existing data (not part of field campaign)	1-2 sites per type	2-3 stations per type/site	using existing data (not part of field campaign)	using existing data (not part of field campaign)	3 replicates per site/station
	Invertebrates (WP4.3)	3 coastal, 3 transitional	using existing data (not part of field campaign)	using existing data (not part of field campaign)	7 stations per type	using existing data (not part of field campaign)	-	3 replicates per station
	Fish (WP4.4)	3 transitional	-	3 sites	3 stations	using existing data (not part of field campaign)	to be defined	3 replicates per station

Figure 19 Draft plan of WISER field sampling campaign for each WP.

References

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

Clarke, R.T. 2000. Uncertainty in estimates of river quality based on RIVPACS. In: Assessing the biological quality of freshwaters: RIVPACS and similar techniques. Wright, J. F., D. W. Sutcliffe & M. T. Furse (eds), Freshwater Biological Association, Ambleside: 39–54.

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

Clarke, R.T., Furse, M.T., Gunn, R.J.M., Winder J.M and Wright, J.F., 2002. Sampling variation in macroinvertebrate data and implications for river quality indices. Freshwater Biology 47: 1735–1751.

Clarke, R.T., Wright, J.F. and Furse, M.T., 2003. RIVPACS models for predicting the expected macroinvertebrate fauna and assessing the ecological quality of rivers. Ecological Modelling 160: 219–233.

Clarke, R. T. and Hering, D., 2006. Errors and uncertainty in bioassessment methods – major results and conclusions from the STAR project and their application using STARBUGS. Hydrobiologia 566: 433–439.

Clarke, R.T., Lorenz, A., Sandin, L., Schmidt-Kloiber, A., Strackbein, J., Kneebone, N.T and Haase, P., 2006a. Effects of sampling and sub-sampling variation using the STAR-AQEM sampling protocol on the precision of macroinvertebrate metrics. Hydrobiologia 566: 441–459.

Clarke, R.T., Davy-Bowker, J., Sandin, L., Friberg, N., Johnson, R.K. and Bis, B., 2006b. Estimates and comparisons of the effects of sampling variation using 'national' macroinvertebrate sampling protocols on the precision of metrics used to assess ecological status. Hydrobiologia 566: 477–503.

Council of the European Communities, 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(43): 1–72.

Devlin, M., Best, M. & Haynes, D. 2007. Special issue Editors: Implementation of the Water Framework Directive in European marine waters. Marine Pollution Bulletin, 55, 3-4.

Ellis, J.E., 2007. Combining multiple quality elements and defining spatial rules for WFD classification. Environment Agency Science Project Number SC060044, Product Code GEHO0807BMXZ-E-E.

Furse, M.T., Clarke, R.T., Winder, J.M., Symes, K.L., Blackburn, J.H., Grieve, N.J. and Gunn, R.J.M., 1995. Biological assessment methods: controlling the quality of biological data. Package 1: The variability of data used for assessing the biological condition of rivers. Bristol: National Rivers Authority, (R&D Note 412).

Moss, D., Furse, M.T., Wright, J.F., Armitage, P.D., 1987 The prediction of the macro-invertebrate fauna of unpolluted running-water sites in Great Britain using environmental data. Freshwater Biol., 17: 41-52.

Moss, D., Wright, J.F., Furse, M.T., Clarke, R.T., 1999. A comparison of alternative techniques for prediction of the fauna of running-water sites in Great Britain. Freshwater Biol., 41: 167-181.Ostermiller, J. D. and Hawkins, C. P., 2004. Effects of sampling error on bioassessments of stream ecosystems: application to RIVPACS-type models. Journal of the North American Benthological Society 23: 363–382.

Wright, J.F, Sutcliffe, D.W. and Furse, M .T., eds, 2000. Assessing the biological quality of fresh waters: RIVPACS and other techniques. Freshwater Biological Association, Ambleside, 373pp.