UKTAG Environmental Standards

GB Lake Fish e-DNA Assessment Procedure

by

Water Framework Directive – United Kingdom Technical Advisory Group (WFD-UKTAG)

Publisher: **Water Framework Directive – United Kingdom Advisory Group (WFD-UKTAG)**

c/o Environment Agency
Horizon House
Deanery Road
Bristol
BS1 5AH
www.wfduk.org

September 2021

ISBN: 978-1-84911-484-4

Health and safety statement

WARNING. Working in or around water is inherently dangerous; persons using this standard should be familiar with normal laboratory and field practice. This published monitoring system does not purport to address all of the safety problems, if any, associated with its use. It is the responsibility of the user to establish appropriate health and safety practices and to ensure compliance with any national regulatory guidelines.

It is also the responsibility of the user if seeking to practice the method outlined here, to gain appropriate permissions for access to water courses and their sampling.

UKTAG Environmental Standards – Lake Fish

1. Introduction

This method statement describes how to determine the WFD class for the lake fish biological element. Lake fish standards have been developed for the first time for use in the third river basin planning cycle.

The lake fish classification tool is used to describe the impact of nutrient pressures on fish populations, by calculating an Ecological Quality Ratio (EQR) value. This describes the similarity of the observed fish community to that expected under reference conditions. The approach uses five metrics, developed from analysis of eDNA obtained from water samples.

The tool differentiates lake fish communities into one of four classes: High, Good, Moderate, and Poor, by applying boundary values to the EQR results. Bad status is provisionally reserved for sites where the fish community is very seriously impacted or no fish eDNA is detected when fish are expected to be present. No such sites were observed in the training dataset used to develop this tool.

2. Lake fish classification tool

2.1 Method summary

Data derived from environmental DNA, or eDNA, has been used to develop a new tool for classifying fish in lakes. It is based on analysis of DNA recovered from water samples. The technical details of the development of the tool are available in Wilby *et al*. (2020).

Research in Britain, funded mainly by SEPA and the EA, has demonstrated that eDNA meta-barcoding provides both qualitative and to some degree quantitative information on fish communities in large lakes, outperforming established methods in terms of the number of species detected.

The lake fish method uses a metric-based approach to generate an EQR value for each lake based on the sample occupancy of different species or combinations of species. Five metrics are used, selected based on their ability to differentiate between sites with high and low nutrient pressure. These are:

- Brown trout
- Percidae
- Roach
- Salmon+charr+coregonids
- Carp+bream

Table 1. *The taxa and metrics used to calculate EQR values for the GB* Lake Fish e-DNA Assessment Procedure.

Common name	**Species name**	**Family**	**DNA Primer Identification level**	**Community Metric Grouping / Species**
Brown trout	*Salmon trutta*	Salmonidae	*Salmon trutta*	Brown trout
Perch	*Perca fluviatilis*	Percidae	*Perca fluviatilis/ Sander lucioperca*	Percidae
Zander*	*Sander lucioperca*	Percidae	*Perca fluviatilis/ Sander lucioperca*	Percidae
Roach	*Rutilus rutilus*	Cyprinidae	*Rutilus rutilus*	Roach
Arctic charr	*Salvelinus alpinus*	Salmonidae	*Salvelinus alpinus*	Salmon+charr+coregonids
Atlantic salmon	*Salmon salar*	Salmonidae	*Salmon salar*	Salmon+charr+coregonids
Powan	*Coregonus lavaretus*	Coregonidae	Coregonus spp.	Salmon+charr+coregonids
Vendace	*Coregonus albula*	Coregonidae	Coregonus spp.	Salmon+charr+coregonids
Common carp	*Cyprinus carpio*	Cyprinidae	*Cyprinus carpio*	Carp+bream
Bream	*Abramis brama*	Cyprinidae	*Abramis brama*	Carp+bream

*Primer used cannot discriminate between Perch and Zander, and the white fish-vendace and powan found in GB lakes.

The metrics are all based on “site-occupancy”, which is the proportion of samples which had a positive result for the presence of the particular fish species (or group of species).

The observed values for these metrics are compared to expected values under low or no pressure conditions. Expected values were developed during the tool development from analysis of key environmental variables (morpho-edaphic index, lake area, altitude, distance from sea) from a subset of good quality reference sites.

The EQR for each metric is then determined from the ratio of Observed/Expected scores, which are normalised to give a value between 0 and 1. The final combination rule to generate the overall fish EQR for each lake is a simple averaging of the five separate normalised metric EQRs.

2.2 Environmental pressures to which the method is sensitive

The overall fish EQR has a highly significant relationship with eutrophication pressures that

is of a comparable strength to the relationship that other BQEs exhibit with eutrophication indicators in lakes (e.g. chlorophyll, macrophytes or TP). Testing during method development using a range of fish metrics found no evidence of sensitivity to hydro-morphology pressures.

3. Data and information requirements

3.1 Environmental sampling

Water sampling should follow the procedure described in Hänfling *et al.* (2016). The following are the key points:

- Samples should be taken from 20 shoreline locations spread approximately equidistantly around the lake perimeter.
- Samples should be taken from shallow water (>50cm depth)
- Sample sites should avoid areas of significant inflow.
- Samples should be taken between December and March.
- Two litres of water should be taken from each site, each comprised of 5 x 400 ml subsamples taken at approximately 5 m intervals along a transect parallel to the shore.
- Suitable care must be taken to avoid contamination during sampling and subsequent sample storage and transportation. A least one sampling blank control must be included from each lake sample batch to ensure that this is achieved.

3.2 eDNA capture and extraction

Water samples must be filtered within 24 hours of collection. The filter membrane must be composed of mixed cellulose ester or another suitable hydrophilic material such as cellulose nitrate (CN), Polyethersulphone (PES), Polycarbonate (PCTE) or other materials demonstrated to perform equivalently. The filter should include a membrane with a pore size of 0.8 µm or less. Ideally, the full 2L sample should be filtered but if this is not possible due to filter clogging, a minimum of 1 litre of water must be filtered. This can be achieved with more than one filter if these can be combined during the DNA extraction.

Filters can be open or in a protective housing (enclosed filters). When open filters are used filtration must be carried out in suitable sterile dedicated facility to minimise contamination and control blanks must be used during filtration. If using enclosed filters water filtration can be carried out in the field. After filtration filters should be preserved immediately (within 30 minutes) through freezing at -20 °C or preservation in a suitable solution. Suitable preservative solutions include molecular grade ethanol (C_2H_5OH) of at least 96% concentration and Longmire's solution (100 mmol/l Tris, 100 mmol/l EDTA, 10 mmol NaCl, 0,5 % (w/v) SDS) but may include other solutions shown to maintain DNA integrity for an appropriate length of time at ambient temperature. Solutions must be certified target-DNA-free or mixed in a cleanroom environment and tested to ensure that they are target-DNA-free prior to use.

DNA should then be extracted from the filters using an appropriate method, such as the Mu-DNA water protocol (Sellers et al. 2018). DNA extraction must be carried out in a suitably sterile, dedicated facility to minimise contamination. Control blanks must be included during the extraction stages. DNA should be eluted in a final volume 100ul of buffer or water. If

different volumes are used the amount of target DNA in the PCR reaction must be adjusted accordingly.

3.3 eDNA metabarcoding

The eDNA metabarcoding must follow the detailed procedure outlined in Annex 1.

3.4 Bioinformatics analysis

The bioinformatics analysis should follow the detailed pipeline described in Annex 2.

The data produced by the bioinformatics pipeline should be summarised as the number of sequence reads per fish species in each sample. A false positive sequence threshold of 0.001 (0.1%) should be used to remove taxonomic assignments that may have resulted from contamination during library preparation or sequencing (De Barba et al. 2014, Hänfling et al. 2016, Port et al. 2016).

Data should then be summarised as the proportion of sampling sites per lake in which a given species was detected (i.e. site occupancy).

3.5 Supporting environmental information

Calculation of the expected values for the metrics requires the morpho-edaphic index (MEI) of each lake to be calculated as an indicator of baseline productivity. This needs:

- **Lake mean depth**, measured in metres. Where there is no bathymetry data and the mean depth is not known accurately, a modelled mean depth may be used, for example derived from a topographic model or from a known maximum depth.
- **Lake mean alkalinity,** measured in meq/L (50mg/L $CaCO_3$ = 1 meq/L).

Data for most lakes is available from the UK Lakes Portal https://eip.ceh.ac.uk/apps/lakes/

If alkalinity data is not available through the UK Lakes portal, it can be generated from 6-12 samples gathered through the year. If it is believed that environmental pressures in the catchment may affect a lake's mean alkalinity value then a mean value for the lake type should be used instead.

The MEI is then calculated as:

$$MEI=\log_{10}(\text{mean alkalinity} / \text{mean depth})$$

4. Calculation of lake fish classification result

The lake fish EQR value is derived by comparing observed sample occupancy values for the five metrics in the focal site to those expected for that site under reference conditions. The morpho-edaphic index (MEI) value for each water body is used in a logistic regression function to generate the predicted values for each metric.

Observed values for each metric should be taken from the sample occupancy results from the bioinformatics analysis. Values for multi-species metrics are additive. Thus they can

range from 0-3 for the Charr+Coergonid+Salmon metric (i.e. 3=all three taxa were present at every sample location in that lake), or 0-2 for the Carp+Bream metric.

To obtain EQR values for each metric, observed values (O) are compared with expected (E). For positive indicators, where a higher value of the metric indicates higher quality (e.g. brown trout occupancy), EQR = O/E is used, whereas for negative indicators, where a higher value of the metric indicates lower quality (e.g. roach occupancy), EQR = (worst – O)/(worst – E) is used.

Prior to being combined each metric raw EQR series is normalised to position values on a cumulative probability frequency curve (thus ranging from 0 to 1) defined by the mean and standard deviation of that series. These values are given in Table 2. The raw EQR values should first be log transformed for the Brown trout, Charr+Coregonid+Salmon, and Percidae metrics, to remove the influence of outliers that were encountered in the training set.

An overall fish EQR is then calculated by taking a simple arithmetic mean of the five individual metric EQR values.

The formulas for calculating each metric are summarised in Table 2, along with the mean and standard deviations for each. These values are calculated automatically in the Lake Fish Calculator, which is available on the UKTAG website http://www.wfduk.org/resources/gb-lake-fish-e-dna-assessment-procedure. Data for sites can be added by over writing the Fishless Lake row, or copying and pasting this row one row down.

Table 2. *Formulae for calculating the expected value of each metric, expressed as site occupancy, and the resultant EQR. The mean and SD values are calculated from the GB dataset used to build the tool, and are used to normalise the EQR value.*

Metric	Expected Occurrence	EQR	Mean	SD
Brown trout	1/(1+(2.103*(7.451^MEI)))	O/E	0.23	0.178
Charr+Coregonid+Salmon	1/(0.33+(2197.544*(18.808^MEI)))	O/E	0.536	0.527
Carp+Bream	0	(2-O)/(2-E)	0.9	0.219
Roach	=1/(1+(73.702*(0.334^MEI)))	(1-O)/(1-E)	0.687	0.417
Percidae	=1/(1+(1.734*(0.074^MEI)))	(1-O)/(1-E)	0.177	0.15

Class boundary values

The boundary values for the classes are given in Table 3 below. These should be used to give a final classification result for each lake. Assuming fish were present values less than 0.24 should be given a class of Poor.

Table 3. *The boundary values used for lake fish classification.*

Class	EQR boundary value
High	0.65
Good	0.42
Moderate	0.24
Poor	<0.24

A fully worked example is given in Annex 3.

5. References

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **2014**:1-7.

De Barba, M., C. Miquel, F. Boyer, C. Mercier, D. Rioux, E. Coissac, and P. Taberlet. 2014. DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. Molecular Ecology Resources **14**:306-323.

Deakin, C. T., J. J. Deakin, S. L. Ginn, P. Young, D. Humphreys, C. M. Suter, I. E. Alexander, and C. V. Hallwirth. 2014. Impact of next-generation sequencing error on analysis of barcoded plasmid libraries of known complexity and sequence. . Nucleic Acids Research **42**:e129.

Edgar, R. C., B. J. Haas, J. C. Clemente, C. Quince, and R. Knight. 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics **27**:2194-2200.

Hänfling, B., L. Lawson Handley, D. S. Read, C. Hahn, J. Li, P. Nichols, R. C. Blackman, A. Oliver, and I. J. Winfield. 2016. Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. Molecular Ecology **25**:3101-3119.

Harper, L. R., L. Lawson-Handley, C. Hahn, N. Boonham, H. C. Rees, K. C. Gough, E. Lewis, I. P. Adams, P. Brotherton, S. Phillips, and B. Hänfling. 2018. Needle in a haystack? A comparison of eDNA metabarcoding and targeted qPCR for detection of the great crested newt (Triturus cristatus). Ecology and Evolution **8**:6330-6341.

Kelly, R. P., J. A. Port, K. M. Yamahara, and L. B. Crowder. 2014. Using Environmental DNA to Census Marine Fishes in a Large Mesocosm. Plos One **9**:e86175.

Kitson, J. J. N., C. Hahn, R. J. Sands, N. A. Straw, D. M. Evans, and D. H. Lunt. 2019. Detecting host–parasitoid interactions in an invasive Lepidopteran using nested tagging DNA metabarcoding. Molecular Ecology **28**:471-483.

Magoč, T., and S. L. Salzberg. 2011. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics **27**:2957-2963.

Port, J. A., J. L. O'Donnell, O. C. Romero-Maraccini, P. R. Leary, S. Y. Litvin, K. J. Nickols, K. M. Yamahara, and R. P. Kelly. 2016. Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. Molecular Ecology **25**:527-541.

Riaz, T., W. Shehzad, A. Viari, F. Pompanon, P. Taberlet, and E. Coissac. 2011. ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. Nucleic Acids Research **39**:e145.

Rognes, T., T. Flouri, B. Nichols, C. Quince, and F. Mahé. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ **4**:e2584.

Schnell, I. B., K. Bohmann, and M. T. P. Gilbert. 2015. Tag jumps illuminated – reducing sequence-to-sample misidentifications in metabarcoding studies. Molecular Ecology Resources **15**:1289-1303.

Sellers, G. S., C. Di Muri, A. Gómez, and B. Hänfling. 2018. Mu-DNA: a modular universal DNA extraction method adaptable for a wide range of sample types. Metabarcoding and Metagenomics **2**: 1-11.

Willby, N,. Law, A,. Bull, C., Bernd Hänfling, B., Lawson Handley, L. and Winfield, I.. 2020. A tool for classifying the ecological status of lake fish in Britain based on eDNA metabarcoding 2020. Report to Scottish Environment Protection Agency.

Zhang, Z., S. Schwartz, L. Wagner, and W. Miller. 2000. A Greedy Algorithm for Aligning DNA Sequences. Journal of Computational Biology **7**:203-214.

Annex 1. eDNA metabarcoding

Libraries should be prepared for sequencing using a nested metabarcoding workflow with a twostep PCR protocol, where Multiplex Identification (MID) tags (unique 8-nucleotide sequences) are included in the **first and second** PCR for sample identification (Kitson et al. 2019).

Dedicated rooms should be used for pre-PCR and post-PCR processes. For the first PCR, lakes should be processed individually to minimise cross-contamination risk and each lake should have unique 2nd PCR MID tags. Controls should be PCR-amplified alongside eDNA samples to screen for sources of potential cross-contamination between lakes. Consequently, the first PCR for each lake consists of 20 eDNA samples, a sampling/filtration blank, an extraction blank, a PCR negative control, and a PCR positive control.

DNA extracts should be PCR-amplified using vertebrate-specific primers that target a 106 bp fragment of the mitochondrial 12S ribosomal RNA (rRNA) region in fish (Riaz et al. 2011, Kelly et al. 2014). These have been validated in vitro for 22 common UK freshwater fishes and in situ on three deep lakes in the English Lake District, and demonstrated their suitability for eDNA metabarcoding of UK lake fish communities (Hänfling *et al.,* 2016). The primers should be modified to include MID tags, heterogeneity spacers, sequencing primers, and pre-adapters.

24 unique MID tags for the forward and 24 unique MID tags for the reverse primers should be included to allow 24 samples (i.e. all samples from one lake plus controls) to each be labelled with a unique forward and a unique reverse primer (i.e. not simply unique combinations) to reduce barcode misassignment and tag jumps (Deakin et al. 2014, Schnell et al. 2015). The same MID tags can be used for all lakes if these are processed separately but all samples from one lake should receive unique 2nd PCR MID tags at the forward and reverse primer.

The first PCR should be performed in triplicate for each sample/control to combat stochasticity arising from low target DNA concentrations (PCR replicates for each sample/control should have the same tag combination). Eight-strip PCR tubes with individually attached lids should be used instead of 96-well plates to further minimise cross-contamination risk between samples (Port et al. 2016). PCR reactions should be performed in 25 µl volumes using 2 µl of template DNA, a High-Fidelity TAQ and include Bovine Serum Albumin in order to reduce PCR inhibition.

The following PCR protocol is recommended but other reagents might be suitable provided they can be shown to give equivalent results: A reaction consisting of 12.5 µl of Q5® High-Fidelity 2x Master Mix (New England Biolabs® Inc., MA, USA), 0.5 µl of Thermo Scientific™ Bovine Serum Albumin (Fisher Scientific UK Ltd, UK), 7 µl of MGW (Fisher Scientific UK Ltd, UK), 1.5 µL of each 10 µM tagged primer (final concentration 0.75 µM; Integrated DNA Technologies, Belgium), and 2 µl of template DNA. PCR reactions should be sealed with mineral oil (Sigma-Aldrich Company Ltd, UK) droplets (Harper et al. 2018).

The following thermocycling profile is recommended but slight variations (**apart from cycle**

number) are acceptable: 98 °C for 5 mins, 35 cycles of 98 °C for 10 s, 58 °C for 20 s and 72 °C for 30 s, followed by a final elongation step at 72 °C for 7 mins.

PCR triplicates should be pooled and a subsample should be visualised using appropriate electrophoretic and staining techniques to compare the relative amplification strength.

All PCR products from one lake should then be pooled according to band strength (no/very faint band = 20 µl, faint band= 15 µl, bright band = 10 µl, very bright band = 5 µl) on gel (Alberdi et al. 2018) to create sublibraries for purification (i.e. removal of non-specific amplicons). A recommended approach is the use of Mag-BIND® RxnPure Plus magnetic beads (Omega Bio-tek Inc,GA, USA), following the double DNA size selection protocol established by Bronner et al. (2009). Ratios of 0.9x and 0.15x magnetic beads to 100 µl of each sub-library should be used. Eluted DNA (25 µl) should be stored at 4 °C until second PCR amplification. Other purification protocols might be suitable, provided they can be shown to give equivalent results.

The second PCR stage binds pre-adapters, MID tags, and Illumina adapters to the purified sublibraries. Sub-libraries should be labelled with a unique forward and a unique reverse primer to reduce barcode misassignment and tag jumps (Deakin et al. 2014; Schnell et al. 2015). Two replicates with the same tag combinations should be performed for each sub-library in 50 µL volumes using 4 µl of template DNA and a thermocycling profile consisting of 10 cycles. The following protocol is recommended:

25 µl of Q5® High-Fidelity 2x Master Mix (New England Biolabs® Inc., MA, USA), 13 µl of MGW (Fisher Scientific UK Ltd, UK), 3 µl of each 10 µM tagged primer (final concentration 0.6 µM; Integrated DNA Technologies, Belgium), and 4 µl of template DNA. PCR to be performed with the following thermocycling profile: 95 °C for 3 mins, 10 cycles of 98 °C for 20 s and 72 °C for 1 min, followed by a final elongation step at 72 °C for 5 mins. Strip tubes should be transported in gripseal bags to a pre-PCR room where primers are added in a UV and bleach sterilised laminar hood.

Duplicates for each sub-library should be pooled, and purified to remove non-specific amplicons. The following protocol is recommended but other approaches might be suitable provided they can be shown to give equivalent results: Mag-BIND® RxnPure Plus magnetic beads (Omega Bio-tek Inc, GA, USA), following the double DNA size selection protocol established by Bronner et al. (2009). Ratios of 0.7x and 0.15x magnetic beads to 50 µL of each sub-library should be used. Eluted DNA (25 µL).

DNA quantities of sub-libraries should be established using a suitable method such as Qubit™ 3.0 fluorometer and normalised by pooling according to sample size (if variable) and concentration to allow equal sequencing depth for each sample. The pooled library should be purified using the same ratios, volumes, and protocol as the second PCR purification.

It is recommended to check the size of the amplicon pool using an appropriate method such as Agilent 2200 TapeStation and High Sensitivity D1000 ScreenTape (Agilent Technologies, CA, USA) to verify secondary product has been removed successfully and a fragment of the expected size (315 bp) remains.

The library should then be sequenced using Illumina sequencing technology to achieve a sequencing depth of approximately 5-10 million raw reads per sub-library (lake). As guidance, this can be achieved by sequencing 10 sub-libraries on a MiSeq® Reagent Kit v3 (600-cycle) (Illumina Inc., CA, USA). It is recommended to add 10% PhiX Control (to improve clustering during initial sequencing) to the library.

Annex 2: Bioinformatics analysis

Raw sequence reads should be demultiplexed using a custom Python script then processed using the custom script metaBEAT (metaBarcoding and Environmental Analysis Tool)v0.97.11 or an equivalent work flow which includes the same analysis elements (https://github.com/HullUni-bioinformatics/metaBEAT).

The recommended approach and steps are described below but individual elements can be substituted with equivalent approaches. Raw reads should be quality trimmed from the read ends (minimum per base phred score Q30) and across sliding windows (window size 5bp; minimum average phred score Q30) using Trimmomatic v0.32 (Bolger et al. 2014). Reads should be cropped to a maximum length of 110 bp and reads shorter than 90 bp after quality trimming discarded. The first 18 bp of remaining reads should also be removed to ensure no locus primer remains.

Sequence pairs should be merged into single high quality reads using FLASH v1.2.11 (Magoč and Salzberg 2011), provided there is a minimum overlap of 10 bp and no more than 10% mismatch between pairs. Only forward reads should be kept for pairs that cannot be merged.

A final length filter (106 bp ± 20%) should be applied to ensure sequences reflect the expected fragment size (106 bp).

Retained sequences **must** be screened for chimeric sequences against the University of Hull custom reference database for UK fish (Hänfling et al. 2016) using the uchime algorithm (Edgar et al. 2011), as implemented in vsearch v1.1 (Rognes et al. 2016). Redundant sequences should be removed by clustering at 100% identity ('--cluster_fast' option) in vsearch v1.1 (Rognes et al., 2016). Clusters should be considered sequencing error and omitted from further processing if they are represented by fewer than three sequences. Non-redundant sets of query sequences should then compared against the University of Hull UK fish reference database (Hänfling et al. 2016) using BLAST (Zhang et al. 2000). Putative taxonomic identity should be assigned using a lowest common ancestor (LCA) approach based on the top 10% BLAST matches for any query that matches a reference sequence across more than 95% of its length at minimum identity of 100%.

Annex 3. Worked example

This worked example is for the Loch Lomond (South Basin) water body (WBID- 49003), and is also included in the lake fish calculator provided on the UKTAG website- http://www.wfduk.org/resources/gb-lake-fish-e-dna-assessment-procedure. The calculator will automatically calculate each step and generate a classification for any new lake provided that results have been summarised into sample occupancy for each metric grouping and if mean depth and alkalinity data are available. Data for sites can be added by over writing the Fishless Lake row, or copying and pasting this row one row down.

Table 1 Example of final results from bioinformatics pipeline joined to Metric groupings:

DNA Species ID	Sample occupancy	Metric Grouping
Abramis_brama	0.2	Carp+Bream
Anguilla_anguilla	0.85	NA
Barbatula_barbatula	0.2	NA
Coregonus spp.	0.95	Charr+Coregonid+Salmon
Esox_lucius	0.7	NA
Gasterosteus_aculeatus	0.9	NA
Gobio_gobio	0.05	NA
Gymnocephalus_cernua	1	NA
Lampetra_fluviatilis	0.4	NA
Leuciscus_leuciscus	0.3	NA
Perca_fluviatilis	0.7	Percidae
Phoxinus_phoxinus	0.7	NA
Platichthys_flesus	0.1	NA
Pungitius_pungitius	0.75	NA
Rutilus_rutilus	0.85	Roach
Salmo_salar	0.8	Charr+Coregonid+Salmon
Salmo_trutta	0.9	Brown trout
Squalius_cephalus	0.05	NA

The summarised observed metric values based on the sample occupancy of the component taxa in Loch Lomond (S Basin) are given in Table 2 below. The mean depth of Loch Lomond (S basin) is 19.5 m, and mean alkalinity is 0.232 meq/L.

Table 2: *Summary site occupancy results for each metric grouping.*

Metric	Observed sample occupancy
Brown trout	0.9
Charr+Coregonid+Salmon	1.75
Carp+Bream	0.2
Roach	0.85
Percidae	0.7

Step 1: Calculate expected values and raw EQR values for each metric

The formulae in Table 2 should be used to calculate the raw EQR values.

Table 3. *Formulae for calculating raw EQR values from MEI and site occupancy data*

Metric	Expected Occurrence	EQR
Brown trout	1/(1+(2.103*(7.451^MEI)))	O/E
Charr+Coregonid+Salmon	1/(0.33+(2197.544*(18.808^MEI)))	O/E
Carp+Bream	0	(2-O)/(2-E)
Roach	=1/(1+(73.702*(0.334^MEI)))	(1-O)/(1-E)
Percidae	=1/(1+(1.734*(0.074^MEI)))	(1-O)/(1-E)

Example from Loch Lomond (S, Basin).

MEI=$\log_{10}$ (alkalinity /mean depth) = $\log_{10}$ (0.232/19.5) = -1.92

Observed (O) occupancy of brown Trout = 0.9
Expected (E) occupancy of brown trout = 1/(1+(2.103*(7.451^-1.92))) = 0.96
Raw EQR = O/E = 0.9/0.96 = 0.94

Observed (O) occupancy of Charr+Coregonid+Salmon = 1.75
Expected (E) occupancy of Charr+Coregonid+Salmon = 1/(0.33+(2197.544*(18.808^-1.92)))=0.12
Raw EQR = O/E = 1.75/0.12 = 14.14

Observed (O) occupancy of Carp+Bream = 0.2
Expected (E) occupancy of Carp+Bream = 0
Raw EQR = (worst – O)/(worst – E) = (2-0.2)/(2-0) = 0.9

Observed (O) occupancy of roach = 0.85
Expected (E) occupancy of roach =1/(1+(73.702*(0.334^-1.92))) = 0.002
Raw EQR = (worst – O)/(worst – E) = (1-0.85)/(1-0.002) = 0.15

Observed (O) occupancy of percidae = 0.7
Expected (E) occupancy of percidae =1/(1+(1.734*(0.074^-1.92)))= 0.004
Raw EQR = (worst – O)/(worst – E) = (1-0.7)/(1-0.004) = 0.301

Step 2. Normalise each raw EQR value

Each metric raw EQR series must be normalized to position values on a cumulative probability frequency curve (thus ranging from 0 to 1) defined by the mean and SD for each metric calculated from the whole GB data series. These are given in Table 4 below.

The raw EQR values should first be log transformed for the Brown trout, Charr+Coregonid+Salmon, and Percidae metrics, to remove the influence of outliers.

Table 4. *Mean and SD values to be used for normalising the raw EQR values for each metric.*

Metric	Mean	SD
Brown trout	0.23	0.178
Charr+Coregonid+Salmon	0.536	0.527
Carp+Bream	0.9	0.219
Roach	0.687	0.417
Percidae	0.177	0.15

Brown Trout:

- Log transform the raw EQR series = Log (0.94 +1) = 0.286
- Based on the population of log transformed EQR values for brown trout occupancy, the overall mean = 0.230 and SD = 0.178
- Position the log transformed EQR on a normal cumulative distribution (values scaled from 0 to 1)

$$f(x|\mu,\sigma) = \frac{1}{2}\left[1 + erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

Where error term: $\mathrm{erf}(x) = 1\pi\sqrt{}\int x - xe - t2dt$
Where raw EQR (x) = 0.286, overall mean (μ) = 0.230 and SD (σ) = 0.178
Gives normalised EQR = 0.627

Charr+Coregonid+Salmon:

- Log transform the raw EQR series = Log (14.14 +1) = 1.18
- Based on the population of log transformed EQR values for Charr+Coregonid+Salmon occupancy, the overall mean = 0.536 and SD = 0.527
- Position the log transformed EQR on a normal cumulative distribution (values scaled from 0 to 1)

$$f(x|\mu,\sigma) = \frac{1}{2}\left[1 + erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

Where error term: $\mathrm{erf}(x) = 1\pi\sqrt{}\int x - xe - t2dt$
Where raw EQR (x) = 1.18, overall mean (μ) = 0.536 and SD (σ) = 0.527
Gives normalised EQR = 0.889

Carp+Bream

- Based on the population of EQR values for Carp+Bream occupancy, the overall mean = 0.9 and SD = 0.219
- Position the EQR on a normal cumulative distribution (values scaled from 0 to 1)

$$f(x|\mu,\sigma) = \frac{1}{2}\left[1 + erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

Where error term: $\mathrm{erf}(x) = 1\pi\sqrt{}\int x - xe - t2dt$

Where raw EQR (x) 0.2, overall mean (μ) = 0.9 and SD (σ) = 0.219
Gives normalised EQR = 0.5

Roach

- Based on the population of EQR values for Roach occupancy, the overall mean = 0.687 and SD = 0.417
- Position the EQR on a normal cumulative distribution (values scaled from 0 to 1)

$$f(x|\mu,\sigma) = \frac{1}{2}\left[1 + erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

Where error term: $\mathrm{erf}(x) = 1\pi\sqrt{\int} \; x - xe - t2dt$

Where raw EQR (x) = 0.15, overall mean (μ) = 0.687 and SD (σ) = 0.417
Gives normalised EQR = 0.099

Percidae

- Log transform the raw EQR series = Log (0.301 +1) = 0.114
- Based on the population of log transformed EQR values for Percidae occupancy, the overall mean = 0.177 and SD = 0.15
- Position the log transformed EQR on a normal cumulative distribution (values scaled from 0 to 1)

$$f(x|\mu,\sigma) = \frac{1}{2}\left[1 + erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

Where error term: $\mathrm{erf}(x) = 1\pi\sqrt{\int} \; x - xe - t2dt$

Where Log transformed EQR (x) = 0.114, overall mean (μ) = 0.177 and SD (σ) = 0.15
Gives normalised EQR = 0.338

Step 3 calculate an overall EQR

The final overall EQR value is a simple arithmetic mean of the individual metric EQR values.

Loch Lomond (S. Basin) EQR= (0.627+0.889+0.5+0.099+0.338)/5=0.491

Step 4 generate a classification

The boundary values in Table 5 are used to generate a classification result from the overall EQR value.

Table 5. *The boundary values used for lake fish classification.*

Class	EQR boundary value
High	0.65
Good	0.42
Moderate	0.24
Poor	<0.24

Loch Lomond (S Basin) EQR=0.491 which lies between the boundary value for “Good” (0.42), and “High” (0.65) so the overall class for lake fish in Loch Lomond (S. Basin) is **Good Status.**