

## ASSESSING THE QUALITY OF WATER QUALITY ASSESSMENTS: AN ANALYTICAL QUALITY CONTROL PROTOCOL FOR BENTHIC DIATOMS

MARTYN KELLY AND ANNE LEWIS

*(Dr M. G. Kelly, Bowburn Consultancy, 23 Clarence Street,  
Bowburn, Durham DH6 5BB, England,  
[Tel: (+ 44) (0)191 377 2077; Email: [101327,2514@compuserve.com](mailto:101327,2514@compuserve.com)])*  
and

*A. Lewis, Environment Agency, North-East Region, Tyneside House,  
Skinnerburn Road, Newcastle upon Tyne NE4 7AR, England.)*

### Introduction

Analytical Quality Control (AQC) is becoming increasingly recognised as an essential guarantee of the quality of environmental data. Techniques for assessing the quality of chemical data and some biological data (such as chlorophyll concentration) are well developed and, conceptually, quite straightforward. Depending upon the system, either a proportion of the samples are subjected to independent analysis, or samples with known concentrations of a determinand are inserted into routine analytical runs. Matching the "observed" and "expected" values is then a relatively straightforward statistical exercise.

However, where biological methods involve analysis of community structure, an additional source of error, relating to misidentifications of taxa or the failure to find all taxa in a sample, is introduced. This type of analysis is not readily automated and, rightly or wrongly, a failed sample may be seen to reflect directly upon an individual's performance. The development of workable AQC procedures for these situations is both more complicated and requires a greater level of human interaction than might be the case for a chemical determinand. Yet this has to be set into perspective against the growing importance of biological monitoring in water quality assessments and the increasing use made of biological data in prosecutions, in determining water quality zones in rivers, and in building cases for or against expensive capital investment programmes at sewage treatment works.

For these reasons AQC is now beginning to be regarded as an integral part of monitoring networks, and a protocol for assessing the quality of invertebrate survey data (van Dijk 1995) was implemented by the National Rivers Authority (NRA: forerunner of the Environment Agency) in 1995, in time for the quinquennial "General Quality Assessment" survey. At the same time the Trophic Diatom Index (TDI) (Kelly & Whitton 1995), a new tool for the assessment of river eutrophication, was being evaluated by the NRA, and

the opportunity was taken of establishing quality standards for this method right from the start.

Diatoms are particularly amenable to AQC procedures, as standard preparation techniques result in permanent slides that can be readily transported between laboratories. In this article we explain the background to the development of an AQC procedure for the TDI, highlighting some of the statistical and taxonomic problems encountered, and go on to demonstrate how the system works in practice. A fundamental difference between the AQC protocol developed for diatoms, and the one developed for invertebrates, is that most of the people involved in using the TDI in the early stages will be biologists from the NRA and other regulatory organisations who are relative newcomers to diatoms. AQC, and quality assessment procedures in general, have a reputation for generating reams of paperwork and being seen as an unnecessary intrusion into the workings of a laboratory. In the case of the TDI, we hoped that an AQC procedure would be seen as an integral part of the method and part of the "learning curve" for the biologists involved. It was important that the protocol was designed in such a way as to provide both data of a known and defensible quality and constructive feedback for the individuals concerned.

### **Statistical background**

Most diatom-based pollution indices, including the TDI, use changes in the relative proportions of different taxa to indicate changing environmental conditions. The techniques involved are, therefore, much simpler than those involved in many studies of phytoplankton, for example, where absolute numbers are required.

Typically, a diatom analyst takes a small (ca. 100 to 200  $\mu$ l) *random subsample* from the *main sample* of (benthic) diatoms and places it on a microscope slide. Under suitable magnification, the analyst then traverses across the slide, identifying and counting each diatom valve that is seen until a certain number (at least 200 valves) has been counted. The distribution of valves on the slide is normally assumed to follow a Poisson distribution, i.e. the valves are distributed at random. In brief, this means there is a high chance that a particular field of view will contain none or a single individual of a particular taxon, a smaller chance of containing two, a still smaller chance of three, and so on.

The statistical basis of making counts with diatoms was explained by Lund et al. (1958). If the valves are *randomly distributed on the slide*, the statistical error for the recorded count is related to the number of valves that have been counted. The error, expressed as upper and lower confidence limits (usually the 95% limits), can be readily obtained from standard statistical tables for a Poisson distribution. Thus, for example, if the recorded count for a particular

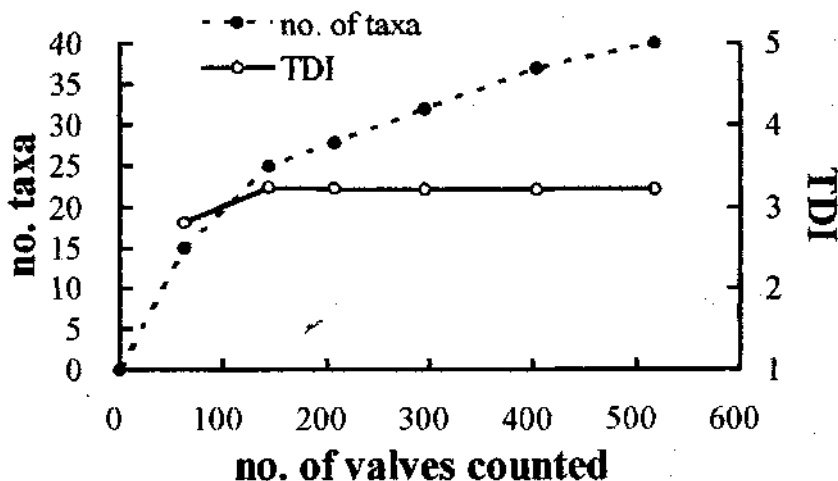


FIG. 1. Relationship between the numbers of taxa observed and the numbers of valves counted, and the Trophic Diatom Index (TDI) for the River Og, a tributary of the River Kennet in southern England. The TDI is calculated from:

$$\text{TDI} = \frac{\sum_{j=1}^n a_j v_j i_j}{\sum_{j=1}^n a_j v_j}$$

where  $a_j$  = abundance (proportion) of species  $j$  in the sample,  $v_j$  = indicator value (1 to 3) and  $i_j$  = pollution sensitivity (1 to 5) of species  $j$ . The value of this version of the TDI can range from 1 (very low nutrient concentrations) to 5 (very high nutrient concentrations). (Reproduced with permission from the *Journal of Applied Phycology*, 7, 433-444.)

taxon is 35 valves, the lower 95% confidence limit for the count is 24 and the upper 95% confidence limit is 49. These are relatively wide limits but accord with the exponential growth patterns of algae, which can lead to a doubling of population size within a few hours.

It is relatively easy to compare the proportions of dominant and common taxa in replicated counts made on diatoms on a single slide. For example, a taxon that comprises 20 valves in a total of 200 counted valves will have 95% confidence limits of 13.2 and 29.1. The number of valves of this particular taxon found during a replicate count on the same slide should therefore fall within these limits. Greater precision could be achieved by counting more valves, but four times the number would have to be counted in order to increase precision by a factor of two (Lund et al. 1958). This, however, involves greater manpower costs with little or no improvement in the resulting

water quality prediction (Kelly & Whitton 1995), due to the manner in which indices such as the TDI are calculated, using weighted average equations (see the legend to Fig. 1). Furthermore, because we are counting diatom valves that are randomly distributed on the microscope slide, occasional and rare taxa may well be missed in a second count made by another analyst or *even by the same individual*. Conversely, a second count might find taxa that were not encountered or recorded in the first count. This problem is particularly acute with taxa whose overall count is less than 4, as the lower 95% confidence limit is then below 1.

### **Taxonomic aspects of quality control**

A neglected aspect of biological AQC is the importance of a clear understanding of what is being counted. This is more than just sound taxonomy: it also involves practical decisions about appropriate taxonomic levels. If there is no prior agreed standard, individuals will identify to whatever level feels most comfortable. One benefit of a standard index such as the Biological Monitoring Working Party score (for invertebrates) or the TDI is that it relies upon a finite number of taxa (86 for the TDI) and the framework of the index itself guides users to the level of taxonomy appropriate to the study.

The TDI was designed in such a way that taxa which are not easily recognised, or for which relatively little environmental data are available, are lumped together. This does not preclude individuals from identifying species within these groups so long as it is made clear to users of the data that this level of taxonomy has not been verified. Although individuals are often very confident about their own taxonomic abilities, the scant evidence available from inter-laboratory comparisons (Munro et al. 1990; Kelly, unpublished data) is that there are often considerable variations between the specific or varietal names applied to particular morphological forms of diatoms. Furthermore, a lot of practical identification relies upon intuition and insight gained through observations over the course of an individual's career, and inter-laboratory comparison exercises can be emotionally-bruising episodes for all concerned. From a purely practical point of view, therefore, it is sensible to let the level of an AQC exercise be set to maximise the information to be extracted, rather than to use all the data that is inputted.

### **Objectives of a protocol for AQC in the enumeration of benthic diatoms**

On the basis of the preceding arguments, three criteria may be established for the design of biological AQC protocols:

- is the exercise capable of detecting significant deviations in accuracy or precision of data (identification and counts of various taxa) produced in this laboratory?;

- does it produce sufficient documentary evidence to persuade outsiders that data produced is of a known and defensible quality?
- have steps been taken to ensure that collection and collation of AQC data is seen as integral to an organisation's mission rather than as another bureaucratic headache?

The principles of enumeration of diatoms for water quality purposes are identical to those of palaeolimnology (Battarbee 1986). The analyst puts a prepared slide on the microscope stage, focuses, finds the edge of a patch of diatoms and then slowly moves the stage horizontally underneath the objective, recording every diatom valve seen. When the other side of the patch of diatoms is reached, the analyst moves down (or up) the slide and starts a new traverse. This is continued until the required number of valves has been counted.

The objectives of the AQC protocol considered here are to provide independent verification of:

- the taxa found in the subsample and
- the proportions of dominant (nominally more than 10%) taxa.

A count of at least 200 valves is recommended for routine purposes (Kelly & Whitton 1995). A lower limit of four valves (2%) was set for inclusion of taxa within the AQC exercise.

### **Protocol for AQC in the enumeration of benthic diatoms (Fig. 2)**

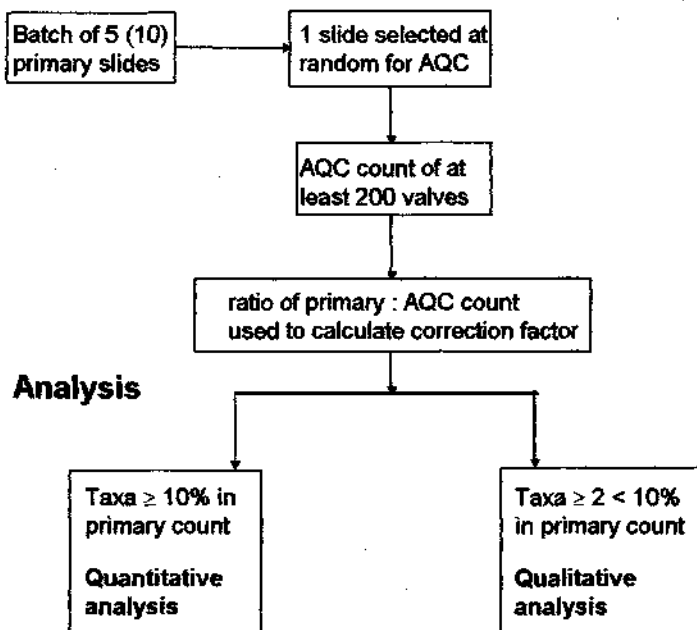
(1) One in ten of all subsamples (prepared slides) per analyst (one in five during the early stages) is *selected at random* and submitted for AQC by an independent analyst. Each set of ten (or five) subsamples represents a "batch" that must be formally linked in some way (by a numbering system, for example) to the subsample submitted for the AQC.

(2) At least 200 valves are counted by the AQC analyst, under identical conditions to those of the primary count. In practice, it is common to go slightly over this total. Numbers in the AQC count therefore have to be adjusted to match the total number counted in the primary count. Thus if the primary count was 213 valves and the AQC count was 221 valves, the correction factor is  $213/221 = 0.96$ .

(3) Quantitative AQC is performed for >10% in the primary count, using 95% confidence limits for a Poisson distribution, taken from tables in standard statistical textbooks.

(4) Qualitative AQC is performed for >2% but <10% in the primary count. All must be recorded at least as "present" in the AQC count.

## Selection



## Action

	Pass	Fail
i.	All taxa $\geq 10\%$ within 95% confidence limits	One or more taxa $\geq 10\%$ outside 95% confidence limits
	and	or
ii.	All taxa $\geq 2\%$ in AQC subsample also found in primary subsample	One or more new taxa found at $\geq 2\%$ in AQC subsample
	and	or
iii.	All taxa $\geq 2\%$ found in primary subsample found in AQC subsample	One or more taxa $\geq 2\%$ found in primary subsample not found in AQC subsample.
	= Pass	= Fail
	Data from batch of subsamples accepted onto database	Batch of subsamples recounted and resubmitted for AQC.

FIG. 2. Flow chart indicating stages in the analytical quality control procedure for subsamples of benthic diatoms. See the text for further details.

- (5) In addition, any taxa 2% in the AQC count must be "present" in the primary count.
- (6) A subsample is considered to have passed AQC if:
- (i) all taxa 10% fall within the confidence limits (step 3)
  - (ii) all taxa 2% in the primary count are present in the AQC count (step 4)
  - (iii) all taxa 2% in the AQC count are present in the primary count (step 5).
- (7) A subsample is considered to have failed AQC if:
- (i) any taxon 10% falls outside the confidence limits (step 3)
  - (ii) one or more taxon fails step 4
  - (iii) one or more taxon fails step 5.
- (8) All subsamples belonging to the batch (of ten) from which a subsample has passed AQC are considered to be verified.
- (9) A subsample that fails AQC is returned to the original laboratory for re-checking. Taxonomic queries may be handled simply by a qualitative check of the slide. Queries regarding step 3, however, may require a partial or complete re-count.
- (10) All subsamples in a batch from which a failed AQC subsample was drawn "must be considered as suspect until the queries have been addressed. This may involve qualitative or quantitative re-examination of some or all subsamples and submission of at least one extra subsample (along with resubmission of the failed subsample) for AQC.

**Worked example: River Blackwater, downstream from Sandhurst sewage works, September 1995**

This example is based upon a genuine subsample, submitted for AQC by one of the NRA laboratories participating in the evaluation of the TDI. The original count is shown in Table 1, along with the appropriate 95% confidence limits (from standard statistical tables). 216 valves were counted in the AQC audit, compared with 200 in the original count, so the first step was to adjust the AQC analysis to represent the number of valves of each taxon that would have been counted if the number in the sample had been 200.

The second most abundant taxon in the sample, for example, is *Achnanthes lanceolata*-type (actually a mixture of *A. lanceolata* and *A. rostrata*). 41 valves were counted in the AQC sample, but this was corrected to 38 in order to make it directly comparable with the original count (see above; protocol

Table 1. Worked examples of an AQC exercise, based on a subsample for the River Blackwater below Sandhurst STW. Original and AQC counts refer to counts made by a NRA biologist and MGK respectively. Lower CL and Upper CL = 95% confidence limits taken from tables, corrected to the nearest integer. N/A = not applicable. The correction factor for the AQC count was  $200/216 = 0.93$ .

Taxon	Original Count			AQC Count		Pass/fail
	No. of valves	Lower CL	Upper CL	No. of valves	Corrected	
<i>Navicula</i> – small	40	28	55	72	67	Fail
<i>Achnanthes lanceolata</i>	30	20	43	41	38	Pass
<i>Cocconeis placentula</i>	24	15	36	39	36	Pass
<i>Achnanthes minutissima</i>	22	13	34	3	3	Fail
<i>Nitzschia</i> – other	22	13	34	3	3	Fail
<i>Navicula</i> – other	21	13	33	1	1	Fail
<i>Cocconeis pediculus</i>	21	13	33	0	0	Fail
<i>Gomphonema</i> – other	5	1	12	0	0	Fail
<i>Navicula lanceolata</i>	4	1	11	3	3	Pass
<i>Diploneis</i>	3	0	9	0	0	N/A
<i>Amphora</i> – other	2	0	7	2	2	N/A
<i>Nitzschia amphibia</i>	2	0	7	20	19	N/A
<i>Nav. cryptotenella</i> -type	1	0	6	20	19	N/A
<i>Gom. olivaceum</i>	1	0	6	0	0	N/A
<i>Cymbella</i> – other	1	0	6	0	0	N/A
<i>Pinnularia</i>	1	0	6	0	0	N/A
Others				12	11	N/A
Total	200	–	–	216	–	–

step 2). This value falls within the confidence limits set for the original count (Table 1).

There were, however, several errors in this sample: not surprising, perhaps, as the biologists performing the work were new to the task. In the process of performing the AQC analysis, a number of possible reasons for these "fails" became apparent. Three are considered below.

(1) The high number of *Achnanthes minutissima* was surprising in a lowland, organically-polluted river. On examination, it was clear there were a number of small *Navicula* species that looked very similar to *A. minutissima* valves. The distinction is a difficult one for a beginner, but the absence of the characteristic "bent" girdle views of *A. minutissima* is one clue. A critical



examination should reveal that all the valves possess raphes, whereas in a population of *A. minutissima*, about half of the valves would be rapheless.

If it was indeed the case that small *Navicula* spp. had been confused with *Achnanthes minutissima*, then we should look at the sum of the two taxa. This (40+22) is very close to the 67 valves of small *Navicula* recorded in the AQC count.

(2) It is possible that some of the *Nitzschia* species have also been mis-identified. 22 valves of "*Nitzschia* - other" (see Table 1) were recorded in the original count, along with just two valves of *Nitzschia amphibia*, whereas in the AQC count, the figures were 3 and 20 respectively.

(3) 21 valves of "*Navicula* - other" were also identified in the original count. The most abundant *Navicula* species recorded in the AQC count was *N. reichardtiana* var. *reichardtiana* (20 valves). This is included in the calculation of the TDI as part of the "*Navicula cryptotenella*" complex.

If the above reasons do explain the differences between the two counts, then all taxa 10% in the subsample, except *Cocconeis pediculus*, will have passed AQC when the taxa are correctly identified.

Counts of two further taxa were  $2 < 10\%$  in the primary subsample and are therefore subjected to a qualitative AQC. Whilst *Navicula lanceolata* was found in the AQC sample, "*Gomphonema* - other" was not. (It should be noted that only *Gomphonema angustatum*, *G. olivaceoides*, *G. olivaceum* and *G. parvulum* are identified to species in the TDI.)

Finally, in order to ensure that no taxa are missed in the original count, the AQC count is checked and any taxon 2% here, but absent from the original count, is recorded. None fall into this category for the Blackwater subsample.

The subsample from Blackwater failed the AQC and was returned to the laboratory to be re-checked. However, in the course of the AQC analysis, several possible reasons were identified and passed on to the laboratory. The whole exercise was conducted in an atmosphere of friendly co-operation and the comments were accepted by the biologists involved as part of the learning process. It is their experience from invertebrate AQC procedures that AQC failures decrease as more confidence and experience is gained - partly through the use of internal and external quality control measures.

As more biologists gain experience of identifying diatoms, external AQC could be replaced by internal systems, with a measure of external audit to measure the effectiveness of the AQC.

## Concluding comments

The emphasis throughout has been to create a "user-friendly" approach to AQC which will both produce data of a consistent quality and provide ongoing training for the NRA (now Environment Agency) staff involved in the exercise. It is our view that AQC is an integral part of any biological

method used in water quality assessment and that it must be seen as essentially a practical management tool rather than a sterile statistical exercise. When running properly, an AQC system should prevent too much time being spent on each subsample, as analysts get regular feedback on their performance. By setting a standard for "acceptable" quality, it reduces the temptation for individuals to exceed this: understandable for an academic scientist but not necessarily cost-effective or efficient for a scientist working in a regulatory organisation.

It has been suggested (Cheeseman & Wilson 1978) that 10 to 20% of the effort devoted to routine analyses is required for quality control and, furthermore, that it is better to obtain 10 to 20% fewer results of a known accuracy rather than more results of unknown accuracy. If this seems to be unacceptably high in these cost-conscious days, it is worth remembering that current management theory draws a clear link between an organisation's performance and the training and development of its staff. We hope that we have made it clear that to invest in quality of biological analyses is, by definition, to invest time and effort in staff development.

### Acknowledgements

The views expressed in this article are not necessarily those of the Environment Agency. We thank Dr D. W. Sutcliffe for useful comments on a draft of the manuscript.

### References

- Battarbee, R. W. (1986). Diatom analysis. In *Handbook of Holocene Palaeoecology and Palaeohydrology* (ed. B. E. Berglund), pp. 517-570. Wiley, Chichester.
- Cheeseman, R. V. & Wilson, A. L. (1978). *Manual on Analytical Quality Control for the Water Industry*. Water Research Centre Technical Report No. 66, Water Research Centre, Medmenham. 157 pp.
- Kelly, M. G. & Whitton, B. A. (1995). The Trophic Diatom Index: a new index for monitoring eutrophication in rivers. *Journal of Applied Phycology*, 7, 433-444.
- Lund, J. W. G., Kipling, C. & Le Cren, E. D. (1958). The inverted microscope method of estimating algal numbers and the statistical basis of estimations by counting. *Hydrobiologia*, 11, 143-170.
- Munro, M. A. R., Kreiser, A. M., Battarbee, R. W. et al. (1990). Diatom quality control and data handling. *Philosophical Transactions of the Royal Society, Series B*, 327, 257-261.
- van Dijk, P. (1994). *Analytical Quality Control for Macroinvertebrate Enumeration*. R & D Note 331, Bristol, National Rivers Authority. 37 pp.