NRA Project 239
(239 (PRS 2274-M))

Code of Practice for Handling Data

A Discussion Document

WRc plc

R&D PRS 2274-M

NRA

National Rivers Authority

CODE OF PRACTICE FOR HANDLING DATA - A DISCUSSION DOCUMENT

Report No: PRS 2274-M

October 1989

Author: J C Ellis

Contract Manager: J C Ellis

Contract No: 4744

Client's Reference No: 6.1.4b

RESTRICTION: This report is restricted to the National Rivers Authority and equivalent bodies in Scotland and Northern Ireland

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

WRc Medmenham, Henley Road, Medmenham, PO Box 16, Marlow, Buckinghamshire SL7 2HD. Telephone: Henley (0491) 571531

SUMMARY

I OBJECTIVES

To set out the proposed elements of a national Code of Practice for handling data.

II REASONS

Different statistical assumptions can produce substantially different results, and so it is important that a consistent, national approach is agreed.

III CONCLUSIONS

The NRA would gain considerable benefit from the adoption of a Code of Practice for handling data. By establishing consistent, objective procedures for the handling and statistical interpretation of water quality data, such a Code would ensure the more effective use of monitoring resources both within and between NRA Regions, and would also help to clarify debate on quality issues between the NRA and other organisations.

IV RECOMMENDATIONS

The two main recommendations are:

* that the NRA approves in principle the development of such a Code of Practice; and

* that a small group of quality officers and data users is established - initially to debate the various questions flagged up in the report, and thereafter to serve as a focal point for discussion as development of the Code of Practice progresses.

V RESUME OF CONTENTS

Following an introductory section, the report falls into two main parts. In Section 2 we demonstrate the need for a data handling Code of Practice, taking three topical illustrations relating to river quality assessment, sewage effluent compliance, and load estimation. We go on to propose a list of 13 topics to be covered initially by the Code.

Then in Section 3 we discuss in turn each of these proposed topics. Illustrations are given of current anomalies or inconsistencies between Regions; and where possible we outline the procedure or technique that we would recommend as the basis of the Code of Practice in that area.

CONTENTS

CONTENTS continued

SECTION 1 - INTRODUCTION

1.1 BACKGROUND

WRc's current research programme for the NRA contains five projects relating to statistical aspects of routine quality monitoring. These are:

4739: Sampling Programme Design

4743: Data Handling and Information Needs
4745: Software for Data Interpretation
4744: Code of Practice for Processing Data

4761: Sampling and Statistics Service

Viewed as a whole, the five projects form a natural progression from WRc's recently published Sampling Handbook (Ellis, 1989) - the emphasis now being very much on implementation. The first project - Sampling Programme Design - covers our specific involvement in a 'technical adviser' capacity with the various working groups set up earlier this year by NRA Central to establish sampling guidelines for effluents, rivers, groundwaters and estuaries. At the other extreme, the Sampling and Statistics Service project is a general NRA helpline offering advice to anyone who has a statistical enquiry or request.

The other projects focus in turn on three aspects identified in the Sampling Handbook as being of key importance to the success of a routine monitoring programme:

i) The first - Data Handling and Information Needs - is concerned with the questions: 'What types of information are we hoping to obtain from our monitoring programme?' and 'What is the best way of extracting that information from the data?' That project forms the subject of WRc report PRS 2273-M - a companion to the present report.

ii) Once those information needs have been quantified, the next requirement is for suitable computer software. Some readers will already have seen WRc's AARDVARK package in action, running on their own micro and giving new insights into their own data. AARDVARK is a forerunner to other software products that we are in the process of developing in our Software for Data Interpretation project - the aim, as always, being to help NRA quality officers extract the maximum information from their data.

iii) Finally, our project on a Code of Practice for Handling Data is concerned with establishing standard protocols for the transfer and statistical interpretation of data; and this is the subject of the present report.

1.2 STATUS OF REPORT

Following preliminary visits to several of the NRA Regions, a short note was circulated flagging up some of the topics to be discussed in subsequent more extensive meetings with the Regions. Seven such visits took place during July and August. One aim of this report, accordingly, is to summarise those parts of the discussion relating to a Code of Practice for processing data.

As originally envisaged early in 1989, the intention was to present a definitive Code of Practice at this stage of this project. For three reasons, however, we believe it is both necessary and advantageous to extend the timescale and regard the present report more as an interim document for discussion:

i) There is a substantial overlap between this project and the activities of various of the working groups set up by NRA Central. Of particular relevance are (i) the 1990 River Quality Survey Group, and (ii) the recently established Consent Compliance Policy Group. In these areas, clearly there would be little purpose in suggesting elements of a Code of Practice in advance of the relevant Groups' recommendations.

ii) The time available for consultation has necessarily been limited during the very busy period leading up to the NRA's September launch. The project would therefore benefit from a breathing space in which we could receive more feedback - via further visits to and discussions with the NRA Regions, perhaps coupled with a workshop-type meeting at Medmenham.

iii) The final point is that - as we will be discussing in Section 3 - a number of the proposed items for inclusion in a Code of Practice require further work and discussion before an authoritative industry standard could in any case reasonably be proposed.

1.3 SCOPE OF REPORT

The report is concerned solely with routine chemical quality monitoring data - though we would hope that the principles of a Code of Practice could later be extended to cover biological and bacteriological data. One-off or ad hoc surveys, though of great importance, also lie outside the present remit. Although the emphasis will vary, much of the discussion applies more or less equally to effluents, rivers, groundwaters and bathing waters.

SECTION 2 - ASPECTS OF A DATA HANDLING CODE OF PRACTICE

2.1 NEED FOR THE CODE OF PRACTICE

Three topical examples will illustrate the need for definitive guidance on the ways in which routine monitoring data is interpreted.

(i) River quality assessment

Suppose we have a set of river quality data stretching back over the last seven or eight years. A number of questions can be posed:

* Which Class is the river in?
* Has river quality changed from the 'presumed' Class (ie the Class reported in the last River Quality Survey)?
* Is river quality showing an improving or deteriorating trend?

All such questions are of key importance to the NRA. But the answers themselves depend on a variety of subsidiary questions of a statistical nature (Is quality log-Normally distributed? What do we do about 'fliers'?... or less-than values? Where do we place the burden of proof in testing 95%iles against Class limits? What misclassification risks can we tolerate?) In the absence of unambiguous, authoritative guidance on these matters, a given set of data could produce substantially different 'conclusions' depending on who was looking at it. Clearly such a state of affairs is undesirable.

(ii) **Sewage effluent compliance**

Many readers will recall the confusion that existed in 1984 over the reporting of sewage effluent compliance. Interim consent standards had been derived on the basis of current practice, and so the confident expectation was that nearly all effluents would pass. And yet the percentages of treatment works failing varied from above 90% in one Authority to below 40% in others. Why? Almost entirely because of differences in the rules used across the industry to interpret effluent quality data - coupled with a failure to appreciate the inevitable effects of statistical sampling error. This confusion would have been entirely avoided had there been a Code of Practice giving a definitive account of the approved method.

The implementation of the Look-up Table in 1985 removed much of that uncertainty. Even so, some confusion remains to the present day (apparently because of a drafting error by DoE) over whether it is sample or determinand compliance that is to be assessed by the Look-up Table. Again, a Code of Practice would have made this and other areas of doubt quite clear from the outset.

(iii) **Estimating loads of Red List substances**

A common situation is for the quality officer to have access to daily flow records, but to have concentration values for only 12 or maybe 24 grab samples over a year's monitoring. To estimate annual load from this starting point presents a real statistical challenge. There are half a dozen possible ways of calculating the estimate; and deciding how to calculate confidence limits around the estimates is a problem in its own right. When we add in the extra complication of values less than the analytical limit of detection - a common occurrence with some Red List substances - load estimation becomes a task that demands the sort of definitive guidance that would be provided by a Code of Practice.

The above examples illustrate how desirable it is, when data from the Regions is collated to produce a national picture - or used to provide between-region comparisons - for the assessments to be made in a coherent and consistent manner. Otherwise, apparent variations could simply be artefacts due to the use of different statistical methods rather than indicative of real effects.

A useful parallel can be drawn here between statistical analysis and chemical analysis. The scope for confusion over the interpretation of a single analytical determination has long been appreciated. Thus the SCA standard methods can be thought of as a Code of Practice setting out approved methods of chemical analysis, whilst on the implementation side, AQC provides a set of objective procedures enabling uncertainties in the laboratory to be quantified and controlled. This analogy offers further support to the idea of a similar set of protocols for statistical analysis.

In this report we are primarily discussing the benefits that the NRA would gain from supporting a Code of Practice for handling data. But there are also wider implications. Many other individuals and organisations have a legitimate interest in water quality data -

including the Utilities, HMIP, DoE, the European Commission, various environmental bodies and the general public. A Code of Practice that was recognised by all parties as providing a set of sound, objective procedures for interpreting quality data would be a very helpful aid to rational debate.

2.2 AIMS OF THE CODE OF PRACTICE

A Code of Practice would have the following main aim:

i) to provide a comprehensive and authoritative guide to approved practice in the transfer, analysis and assessment of water quality data.

The Code would also have two important subsidiary aims:

ii) to promote and encourage the consistent and correct interpretation of water quality data; and

iii) to serve as a central register of experience in the use of relevant methods, and so lead to documentary evidence for improvement where necessary.

As a longer-term aim, it would also be worth considering the possibility of an established Code of Practice forming the basis of an eventual British Standard on the Handling of Water Quality Data.

2.3 PROPOSED CONTENT OF THE CODE OF PRACTICE

One of the issues discussed during our recent meetings with the Regions was the question of the topics to be covered by a Code of Practice, and the proposals put forward in the list below are the outcome of those discussions. But the list is by no means complete; indeed, we hope that the period of feedback following the distribution of this discussion document will allow the list to be amended so as to reflect as closely as possible the interests of all those in the industry concerned with the interpretation of water quality data.

The proposed topics for inclusion are:

i) data validation;

ii) data transfer protocols;

iii) handling less-than values;

iv) identifying outliers;

v) testing goodness of fit of statistical distributions;

vi) estimating percentiles;

vii) reporting summary statistics;

viii) estimating loads;

ix) assessing compliance with RQOs;

x) procedures for 1990 River Quality Survey;

xi) assessing effluent compliance;

xii) detecting trends;

xiii) Test Data Facility.

Each of these topics is discussed in turn in Section 3.

2.4 LAYOUT OF THE CODE OF PRACTICE

For each topic, the Code of Practice would provide the following details:

i) the aim of the method or procedure;

ii) its applicability (for example, the statistical assumptions needed; the relevant types of water quality data);

iii) a theoretical justification of the method (referring if necessary to a standard reference) - perhaps including counter-arguments to refute a faulty rival method!;

iv) a full description of how the method is applied, together with a worked example;

v) FORTRAN-77 computer code for applying the method (perhaps incorporating commercially available subroutines such as the NAg library where necessary) - supplied as hardcopy listing, and also available on diskette.

SECTION 3 - PROPOSED TOPICS FOR INCLUSION IN THE CODE OF PRACTICE

3.1 DATA VALIDATION

The conventional approach used to test the validity of data before its inclusion on the computer archive is to check that each new determinand value lies within preset global minimum and maximum limits. Because of the understandable tedium of chasing up false alarms, the limits are often positioned at such extreme points that they trap nothing but the most blatantly gross errors. Indeed, some archive users routinely ignore this facility altogether when inputting data.

The first aim of the Code of Practice in this area, therefore, should be to tighten up and standardise the rules by which the present system is operated. But the Code could also usefully go beyond the traditional data validation approach. For example, possible developments include:

i) utilising **temporal correlations** by comparing each current value with the previous value; and

ii) utilising **cross-correlations** by comparing each current value with current values of other associated determinands.

The more 'smart' the validation system can be made, the better the system will be able to distinguish genuine anomalies from false alarms, and so the greater will be the integrity of the data base for all subsequent uses.

3.2 DATA TRANSFER PROTOCOLS

There is a growing demand for easy, efficient methods of electronic data transfer* - not only within each NRA Region, but also between Regions, and from NRA to other bodies such as WRc, HMIP and DoE. Increasingly the need will be for data to be transferred from a mainframe computer (typically after an archive extraction) to a microcomputer.

Everyone who has been involved in data transfer between computer systems has experienced that familiar feeling of foreboding upon opening a padded bag and finding within it a floppy disk or (worse) a mag.tape. Will we be able to read it?... If so, what's the format?... What's their convention for less-than?... Are there hidden carriage returns?... And so on. By way of illustration, Table 1 shows fragments of three data sets that we have received at WRc within the last few months from different NRA Regions.

What is needed, accordingly, is an agreed standard format for data interchange. For example, the requirement could be for a standard ASCII file containing:

i) a header line;
ii) a row of up to six determinand titles, in the format...........8X,6A12;
iii) then any number of rows (one per sample), in the format....DD/MM/YY,6F12.4;
iv) and then a 99/99/99 date to end.

---

* Except for very small data sets, data transfer by any means other than electronic is extremely tedious and can waste much time and effort (if, that is, anything at all gets done with the data subsequently).

Table 1: Three recent examples of data transfer formats used by NRA Regions in transferring data to WRc

```
                 6     61   9921    77    241   257   207    211   162   183   172   117
               DEPTH   PH    PH     COND  CAL   MAGNE SOD   POTAS  ALKAL SULPH CHLOR NO3-N
   SAMPLE       OF           INSTR   AT   CIUM  SIUM  IUM   SIUM   INITY ATE   IDE
               SAMPL         UMEN  25C                             M.O.
 DATE   TIME                 TAL          MG/L  MG/L  MG/L  MG/L   MG/L  MG/L  MG/L  MG/L
                M      PH    PH    US/CM  CA    MG    NA    K      CACO3 SO4   CL    N
<------->-->

SPT :- 01901ASH-FNL      NGR :- SJ 50514 68883    ASHTON BORE HOLE FINAL   WATER
160919760002       7.65          780                                        156
               SAMPLERS COMMENT- CHLORINATED WATER      ANALYSTS COMMENT-
170919760002       7.5           775                                        157.5
               SAMPLERS COMMENT- CHLORINATED WATER      ANALYSTS COMMENT-
200919760002       7.65          775                                        158.5
               SAMPLERS COMMENT- CHLORINATED WATER      ANALYSTS COMMENT-
```

```
1703140MAGPIE SOUGH OUTFALL  GVCN D35/7             208
17031407506041620AC 15  000000000140000
17031407506041620AC 171 000000000001000
17031407506041620AC 191 000000000001000
17031407506041620AC 20  000000000024000
17031407506041620AC 22  000000000000100
17031407506041620AC 241 000000000000100
17031407506041620AC 41  000000000074000
17031407506041620AC 42  000000000008000
17031407506041620AC 50  000000000280000
17031407506041620AC 54  000000004700000
17031407506041620AC 69  000000000006000
17031407509081440AC 15  000000000190000
```

SAMPLE DATE	TIME	135 SOLIDS SUSPENDED 105C MG/L	85 BOD 5 ATU MG/L O	68 TURBIDITY FTU	111 NITROGEN AMMONIACAL MG/L N
19/05/80	1245	M	2.50	1.80	0.04
21/07/80	0730	M	1.30	1.20	0.14
15/09/80	0915	M	1.30	1.30	0.10
17/11/80	1045	M	1.50	1.10	0.14
19/01/81	0940	M	2.00	1.60	0.12
16/03/81	1005	M	1.90	2.60	0.13
18/05/81	0915	M	1.90	3.90	0.10
20/07/81	1015	M	1.50	1.20	0.09
21/09/81	1015	M	1.60	1.00	0.11
16/11/81	1000	M	2.00	4.10	0.11
18/01/82	1020	M	2.30	2.30	0.07
15/03/82	0945	M	2.10	2.10	0.04
17/05/82	1045	M	3.30	1.80	0.09
04/06/82	1020	M	1.40	1.60	0.06
05/07/82	1410	4.80	1.00	1.75	0.01
19/07/82	0930	M	1.40	1.70	0.05
20/07/82	0935	4.50	1.45	1.00	0.05
03/08/82	1410	6.20	1.20	2.10	0.04

It should not be difficult for experienced archive users to reach consensus upon an appropriate protocol. Indeed, one of the existing National Water Quality Archive retrieval formats may well provide a suitable starting point. The agreed format would then be written into the Code of Practice as the 'default' industry standard. This would not of course preclude the use of more specialised transfer protocols where these were necessary or desirable; but for run-of-the-mill routine applications it would remove a great deal of the hassle currently associated with electronic data transfer.

3.3 HANDLING LESS-THAN VALUES

As we expected, our discussions with the Regions confirmed that a variety of procedures are in use regarding the treatment of less-than values. For each Region, Table 2 below outlines the archive reporting convention used, and the substitution rule(s) used in mainframe calculations involving less-than data.

Plainly it is essential for the NRA to have a consistent policy on less-than values. The ideal arrangement (as we discuss in Appendix 4B of the Sampling Handbook) would be for the actual analytical result to be entered on the archive rather than "<L" or "<C". It would of course need to be flagged appropriately to indicate that the value was not significantly different from zero. But whether or not it was statistically significant, that value would still be the analyst's best estimate of the true but unknown concentration in the sample, and so the calculation could go ahead as normal.

Unfortunately that option is probably not practicable. But for the moment just imagine what those actual values would look like if we did have them to hand. The maximum possible value would be C (because anything larger than C would appear in its own right on the archive). The minimum value could, technically, be negative, but would in practice be balanced by at least as many values on the positive side of zero; so we can assume a minimum value of zero.

Table 2 - Current practice in the Regions regarding less-than values

Value reported as < in archive (C or L)	Substitution rule used (0-----------L)	Comments
L	L	
L	L	
L	L	
L	2L/3	
L	2L/3	
L	L/2	
L	L/2	
C	C	(but archive says it uses C/2!)
C	2C/3	
L	0 or L	(most software uses 0)

Notes: 1. C = Criterion of Detection. (An analytical result must be at least as large as C before the analyst can reliably claim to have detected the substance in question.)

2. L = Limit of Detection = 2C. (If the true concentration is as large as L, the analyst has a 95% chance of obtaining a determination at least as large as C, and so correctly claiming to have detected the substance.)

Now we can propose our suggestion. It is that the calculation be carried out twice - first replacing all less-than values by zero, and then by C. This will give two answers - an optimistic figure and a pessimistic figure - which will automatically bracket the result that would have been obtained had all the data been uncensored.

We believe such a protocol would be preferable to the single application of any sort of substitution rule, for the following reason. If the less-than values are few in number, or are swamped by the magnitudes of the other, uncensored data values, the optimistic and pessimistic values will be numerically very similar and so it will scarcely matter which figure is used. But if there is a wide disparity between the two answers, this is an important conclusion which ought not to be divorced from any subsequent use of the figures. The trouble with reporting just a single answer - a 'point estimate', in statistical jargon - is that however carefully it is hedged around with warning flags and asterisks, people inevitably lift the result out of its context and leave the cautions and provisos behind.

One example of where this suggested protocol would clarify the reporting procedure is in the situation cited by one Region in which all values are below C, and yet C is numerically greater than the EQS! To report that the compliance lay somewhere between 0% and 100%, though at first sight unhelpful, would at least be an unequivocal statement that the analytical method was inadequate for the required task.

Throughout the discussion we have been assuming that all Regions do actually subscribe to the analyst's standard definitions of the Criterion and Limit of Detection (see, for example, Section 8.2.3 of the Sampling Handbook). In practice, this is not always so. One Region indicated, for example, that the standard definitions were followed only when reporting Harmonised Monitoring data; for other purposes a different, more pragmatic convention was used. Several other Regions also expressed doubts about the precise origins of their reported less-than values. This is important because the purpose of the Code of Practice is not just to develop protocols for the future; it must also attempt to disentangle any historical practices which have jeopardised the integrity of existing data on the archive. This is another aspect, therefore, that the Code will need to address.

3.4 IDENTIFYING OUTLIERS

The presence of outliers - data values that appear suspiciously extreme in relation to the main body of data - can seriously distort most types of conclusion drawn from a monitoring programme. They can inflate estimates of the standard deviation; they can bias the mean. Even with non-parametric objectives, such as the judging of compliance with a standard, just one outlier can tip the balance between passing and failing - especially with an absolute or MAC-type standard.

There are clear advantages, therefore, in having some capability to spot potential outliers; and this is another area in which a Code of Practice could offer guidance. Of course, it is one thing to be able to identify a data value as possibly being suspect; it is quite another to be able to establish a sound reason (sample contamination; mix-up in the laboratory; data coding error) for its being discarded. Time is an important factor here: the sooner that possible errors in the sampling and analysis protocol can be flagged up, the greater is the chance of tracking down the genuine mistakes. Thus there are close links here with the possible mechanisms for data validation discussed in Section 3.1.

3.5 TESTING GOODNESS OF FIT OF STATISTICAL DISTRIBUTIONS

Huge amounts of time and energy have been spent over the years debating the types of statistical distribution followed - or not followed - by various river and effluent quality determinands. But with so much of the evidence being anecdotal or of limited applicability, there is still little firm guidance available as to what distributions may be assumed in what circumstances. A good example arose very recently when NRA's Consent Compliance Policy Group was considering how the choice of percentile standard affected the achievable degree of protection against deterioration. The model under discussion made the assumption that if effluent quality were to deteriorate, it would do so in such a way that the coefficient of variation (= standard deviation/mean) remained constant. Although this is a standard assumption made in numerous industry applications, the absence of any authoritative justification prompted the Group to set in train a special study to investigate the assumption further.

With so many thousands of sampling points at issue, a better understanding of the nature of river and effluent quality variability can come only when individual Regions have the capability to test a particular hypothesis on their own data sets. The Code of Practice would help here in three ways. First, it would set out an agreed collection of suitably general 'candidate' distribution models (log-Normal; four-parameter Johnson; gamma; and so on). Next, it would specify an appropriate statistical goodness-of-fit test. Finally, it would offer the software building blocks that would enable each Region to apply a common methodology to its own data.

To the best of our knowledge, only one of the Regions at present has a distribution-testing system in routine use on its archive. This operates on all sewage effluent data sets with 20 or more values, and tests for both Normality and log-Normality using Filliben's Probability Plot Correlation Coefficient test*. It would be useful to build on this Region's experience in developing a wider range of options.

3.6 ESTIMATING PERCENTILES

Much of the debate about statistical distributions mentioned in the previous section was sparked off by the interim consent-setting exercise in the late 1970s, which required the estimation of 95%iles from a great variety of data sets - some rather sparse. With 19 or more data values, a non-parametric method (that is, one making no assumption about the shape of the underlying distribution) can be used to estimate the 95%ile. The attraction of using a parametric method, however, is that a more precise estimate is obtained - provided the assumption is justified.

Table 3 gives some idea of how both the 95%ile estimate and the confidence interval around that estimate can vary depending on the method used. The data - listed in ranked order at the top of the table - consists of 62 suspended solids concentrations, and has been taken from a river quality data set kindly supplied by one of the Regions.

---

* This, incidentally, is one of the tests available in WRc's AARDVARK package.

Table 3 - Illustration of different 95%ile estimation methods for a river quality data set

Ranked suspended solids data:

1.40	1.80	2.00	2.13	2.30	2.40	2.50	2.50	2.60	3.00	3.10
3.10	3.10	3.30	3.40	3.70	3.90	3.90	4.00	4.40	4.50	4.50
4.78	4.80	5.40	5.50	5.60	5.80	6.00	6.20	6.30	6.30	6.60
6.60	6.80	7.00	7.20	7.30	7.40	7.40	7.60	7.60	7.80	8.10
8.20	8.40	8.60	8.80	9.60	9.70	10.60	11.30	11.50	11.80	12.30
12.60	15.90	16.40	16.90	17.80	18.70	37.40				

	90% conf.limits		
95%ile estimation method	Lower limit	Central estimate	Upper limit
Method used by this Region assuming log-Normality(1)	13.1	17.5	23.3
'Exact' method assuming log-Normality	14.5	17.5	22.3
Method assuming Johnson's 4-parameter model(2)	16.1	18.5	21.3
Non-parametric method (Weibull convention)	12.6	17.7	37.4

Note: 1. Because this particular Region happens to use a confidence level of 95%, we have converted its quoted confidence limits to the equivalent values for 90% confidence.

2. The software used by the one Region which assumes the Johnson model contains a statistical error (whose effects we reproduce here) which leads to optimistically narrow confidence limits.

In at least one Region, the non-parametric 95%ile calculation written into its archive software was based on the advice given in the 1970s by the Working Party on Consent Conditions for Effluent Discharges. This recommended that the 95%ile be estimated by the maximum value if the number of samples (n) was four or fewer, or by the second-largest value if n was between four and 20. Unfortunately this produces an optimistically biased estimate of the 95%ile. Table 4 below shows that the actual percentile estimated by the stated procedure ranges from as low as the 67%ile (when n = 5) up to the 90%ile (when n = 19).

Table 4 - Percentiles estimated by various order statistics

No. of samples :	3	4	5	6	7	8	9	10	11	13	15	17	19
Ordered sample value :	3	4	4	5	6	7	8	9	10	12	14	16	18
Percentile estimated :	75	80	67	71	75	78	80	81	83	86	88	89	90

Note: The table follows the Weibull convention, whereby the r-th sample out of n gives an estimate of the 100r/(n+1)%ile.

For parametric estimation, one Region uses Johnson's four-parameter Normality transformation model to produce 95%ile estimates. At present, as we indicate in the footnote to Table 3, there is a problem with the program used to determine confidence intervals around the percentile estimates. The greater flexibility offered by this system is nevertheless a considerable attraction, and its merits deserve further examination. As part of our Sampling and Statistics Service project we are currently engaged in such a study, and we will be reporting on this, along with our other recommendations for 95%ile estimation, later in the year.

3.7 REPORTING SUMMARY STATISTICS

Every Region's archive system can provide a statistical summary of any specified set of data. What is contained in the summary, however, varies widely between Regions. At one extreme, the 'usual' statistics are provided - the minimum, mean, standard deviation, 95%ile and maximum - but with no indication of the assumptions on which the 95%ile was determined, and no confidence intervals accompanying the mean, standard deviation or 95%ile. As one user put it, 'The notion of the confidence interval is an alien concept to the archive.' It was encouraging, nevertheless, to see that summary routines in the majority of Regions do provide confidence intervals.

On the question of 95%iles, a common stance is for routines to assume log-Normality and perhaps also Normality, although the policy in several Regions is to use a non-parametric method wherever possible. In one Region the routine provides all three estimates: Normal, log-Normal and non-parametric. This is an excellent arrangement from the point of view of building up an understanding of how much difference the methods actually make in practice.

Our provisional recommendations for the Code of Practice are these:

i) The statistics should include the minimum, 5%ile, 20%ile, median, mean, 80%ile, 95%ile, maximum, overall standard deviation, 'SDD' estimate of standard deviation*, coefficient of variation, number of values, and number of less-than values.

ii) The preferred method of estimating percentiles should be non-parametric rather than parametric.

iii) Where parametric percentile estimates are quoted, the output should state (a) the assumed distributional model, and (b) whether there is direct supporting evidence.

iv) Approximate 90% confidence limits should be given with every statistic (except for the minimum and maximum).

v) The output should indicate the percentiles estimated by the minimum and maximum. (For example, with $n = 32$ these would be the 3%ile and the 97%ile.)

---

* The SDD statistic is a measure of the variability due to short-term sampling and analytical error. A comparison of the SDD value with the conventional standard deviation gives a useful indication of how much systematic variation there is in the data.

vi) Where the data contains less-than values, separate 'optimistic' and 'pessimistic' summaries should be calculated (see Section 3.3).

3.8 ESTIMATING LOADS

We have already indicated in Section 2.1 the pressing need for a Code of Practice in this area. Because of the particularly knotty statistical and operational problems that arise, we recommended in the Sampling Handbook that a small specialist group be set up to review load estimation methods generally; we reiterate that recommendation here. A good starting point for such a working group would be WRc's recent report (Harrison, Thorogood and Lacey, 1989) describing a study conducted by Essex University on behalf of WRc and DoE to compare load estimation from grab samples and continuous flow-proportional sampling.

3.9 ASSESSING COMPLIANCE WITH RQOS

This is another area of routine quality data interpretation that was identified in Section 2.1 as raising a number of statistical issues requiring careful attention. Some of the content of a Code of Practice on the assessing of compliance with RQOs will follow naturally from the formal mechanisms to be recommended by the 1990 River Quality Survey Group (see Section 3.10 following). Other matters on which guidance would also be desirable would include the procedures to be followed in the years between the quinquennial surveys, and possible methods for detecting changes through time in the percentage compliance with Class limits.

3.10 PROCEDURES FOR 1990 RIVER QUALITY SURVEY

As a special subset of the general advice that would be provided under Section 3.9, there is clearly a need for comprehensive, unambiguous instructions on how the Regions are to handle and interpret the data collected for the 1990 River Quality Survey. NRA's 1990 River Quality Survey Group will of course be paying due attention to these aspects of the survey and advising the Regions accordingly. It would perhaps help to reinforce the Group's recommendations, however, if additionally they were included in the Code of Practice.

3.11 ASSESSING EFFLUENT COMPLIANCE

The NRA has established a Group which is currently reviewing the setting of discharge consents and the way in which compliance with those consents is assessed and monitored. The implications for data interpretation will not be known until some time after the Group has reported to the NRA Board. As we indicated in Section 2.1, however, the need for a Code of Practice in this area of data interpretation is plain whatever the details of the eventual compliance-testing regime should happen to be.

3.12 DETECTING TRENDS

Trend detection is an enormous field of study. In the Sampling Handbook, for example, the discussion on methods for detecting temporal and spatial trends, though far from comprehensive, runs to 60-odd pages. Indeed, the very meaning of the word 'trend' is itself a subject for debate.

In view of the great variety of techniques available, it might be thought that a Code of Practice for trend detection would be unhelpfully restrictive. Historically, however, the industry has always been particularly weak in this area - in part a legacy of the disproportionate attention that has been paid to compliance-related matters in recent years. And in our recent discussions with the Regions, it became evident that there was still a widespread shortage of generally accessible software directed expressly towards the detection of trends in water quality. We suggest, therefore, that a Code of Practice would offer a practical means of helping to remedy this deficiency.

Specifically, we see the Code of Practice - initially, at least - as offering data users a 'starter pack': a basic minimum set of routines of proven value in detecting various types of trend of relevance to water quality management. Subsequently the Regions would be encouraged to try out other potentially useful procedures. As experience in their application was built up, the more promising techniques could be handed on to other Regions and in due course approved for inclusion in the Code of Practice. This process of evolution would ensure the continuing relevance of the Code of Practice to the changing needs of the industry.

3.13 TEST DATA FACILITY

A key requirement for getting the best out of data is to open up a route by which statistical test procedures or 'what if?' routines can easily be tried out on real historical water quality data sets. This can be achieved in one of two ways: by moving the data to the technique, or by moving the technique to the data. We discussed the first of these options in Section 3.2. And certainly a standard protocol would greatly facilitate the transfer of blocks of data to, say, a centre of technical expertise at one of the Regions where they could be submitted to a specialist program.

That approach does, however, have the disadvantage of requiring the transfer and duplicate storage, albeit on a temporary basis, of possibly very large quantities of data. Also, questions of security or confidentiality may arise when data was being passed outside the organisation. For these reasons, there are attractions in the alternative approach of being able to send a piece of software to a particular Region's quality archive and to ask for that software to be applied to various selected data sets. We have coined the term 'Test Data Facility' to describe the system that the Region would need to establish in order to carry out this proposed function.

From the point of view of an individual Region, the Test Data Facility would operate as follows:

i) Receive a request from, say, an NRA Sampling Group to try out a particular FORTRAN subroutine on a selection of data sets.

ii) Extract data for the required timespan from the chosen sampling points and write this to a file.

iii) Take the supplied FORTRAN subroutine and slot this into the standard assessment program previously written by the Region.

iv) Run the assessment program: this will pass each set of data in turn through the routine, and write the resulting output to a file.

v) Transfer the output file to diskette and return this to the instigator of the test exercise.

The water industry does have some experience of this type of exercise. The first attempt was made in 1987, when the WAA River Quality Monitoring and Assessment Group ('the Tyson Group') circulated a proposed algorithm for detecting changes in 95%ile river quality. A number of Regions also contributed in 1988 to a similar collaborative exercise orchestrated by the Mance Group (the precursor to the present NRA River Quality Sampling Group). Still more recently, an exercise was conducted very successfully with Anglian and Severn Trent (perhaps the two Regions with the most experience in this area): this involved the distribution of WRc's EARWIG program to explore the practical consequences of a receiving-water-based approach to the determining of effluent sampling frequencies.

In our current round of discussions with the Regions, the general idea of building on this experience to establish Test Data Facilities across the NRA was received with enthusiasm. Some practical difficulties were foreseen, however, with at least half the Regions declaring that it would be very difficult with their present level of computer expertise to organise such a system. Nevertheless, progress on many of the issues discussed in the report hinges upon the NRA having the ability to test out various statistical assumptions and models quickly and easily across a thoroughly representative selection of data sets. For this reason we believe it important that a trial distribution of software to the Regions be undertaken as soon as possible. Not only will this establish just what the technical computing problems are, it will also bring the benefit of the information derived from the routine itself. We are currently considering what function the trial software could most usefully perform - a cusum-based trend-detection routine is the most likely candidate - and we propose to set the exercise going within the next two months.

SECTION 4 - CONCLUSIONS AND RECOMMENDATIONS

This discussion document has shown the considerable benefits that the NRA would gain from the adoption of a Code of Practice setting out sound, objective procedures for the handling and statistical interpretation of water quality data. Such a Code would ensure the more effective use of monitoring resources both within and between NRA Regions, and would also help to clarify debate on quality issues between the NRA and other organisations.

There are two main recommendations:

* that the NRA approves in principle the development of such a Code of Practice; and

* that a small group of quality officers and data users is established - initially to debate the various questions flagged up in the report, and thereafter to serve as a focal point for discussion as development of the Code of Practice progresses.

ACKNOWLEDGEMENT

Many of the ideas discussed here have their origins in a stimulating discussion paper entitled 'Water Quality Data Analysis - Is there a Role for a "Code of Practice"?, which was presented to the 'WISHART' group in 1988 by John Martin of Severn Trent Water.

REFERENCES

ELLIS J C (1989) Handbook on the Design and Interpretation of Monitoring Programmes. NS29, Water Research Centre.

ELLIS J C (1989) Interim Review of Data Handling and Information Needs of Regulatory Agencies. PRS 2273-M, Water Research Centre.

HARRISON R M, THOROGOOD G A and LACEY R F (1989) Comparison of Load Estimation from Grab Samples and Continuous Flow-proportional Sampling. PRU 2220-M, University of Essex and Water Research Centre.