Water Quality Monitoring - A Practical Guide to the Design and Implementation of Freshwater Quality Studies and Monitoring Programmes
Edited by Jamie Bartram and Richard Ballance
Published on behalf of United Nations Environment Programme and the World Health Organization

ISBN 0 419 22320 7 (Hbk) 0 419 21730 4 (Pbk)

Chapter 14 - USE AND REPORTING OF MONITORING DATA

The preparation of this chapter was co-ordinated by A. Steel with contributions from from M. Clarke, P. Whitfield and others.

The whole system of water quality monitoring is aimed at the generation of reliable data, i.e. data that accurately reflect the actual status of the variables which influence water quality. It is acknowledged that simply generating good data is not enough to meet objectives. The data must be processed and presented in a manner that aids understanding of the spatial and temporal patterns in water quality, taking into consideration the natural processes and characteristics of a water body, and that allows the impact of human activities to be understood and the consequences of management action to be predicted. This is not to say that water quality information must be presented in a way which requires the user to appreciate the full complexity of aquatic systems. The information should provide the user with the understanding necessary to meet the objectives behind the monitoring programme. The importance of setting objectives in the design and implementation of monitoring programmes is discussed in Chapter 3. Objectives imply that the activity has a purpose that is external to the monitoring system; e.g. a management, environmental policy, public health or research purpose. The intent is to use the information to explain water quality, to communicate the information more widely, or to control water quality. Consequently, there is little point in undertaking a monitoring programme unless the resultant data are to be used in fulfilling the objectives.

For some information-users, complex graphical or statistical analysis may be essential, such as for long-term management planning and pollution control in a lake system. For other users, a single variable statistic may be adequate for management purposes. For example, BOD measurement (see Chapter 7) or a water quality or biotic index (see Chapter 11) could be used to evaluate the effectiveness of treatment of domestic or agricultural wastes discharged into a river or stream.

This chapter provides a brief introduction to some of the principal approaches to use and reporting of monitoring data. Full treatment of data handling and presentation for assessments, including details and worked examples of some key statistical methods, is covered in the companion guidebook *Water Quality Assessments*.

14.1 Quality assurance of data

Quality assurance of data is an important precursor to data analysis and use. Effective quality control procedures during sampling and analyses help to eliminate sources of error in the data. However, a second series of data checks and precautions should be carried out to identify any problems that might lead to incorrect conclusions and costly mistakes in management or decision-making. The procedures involved commonly rely on the identification of outlying values (values that fall outside the usual distribution), and procedures such as ensuring that data fall within the limits of detection of a particular method of measurement and checking that normal ion ratios are present. Any anomalous values should be checked because there could be a problem with calculations or with analyses, in which case remedial action should be taken. Genuine outlying values do occur, however, and may be important indicators of changes in water quality.

14.1.1 Basic data checks

Basic data checks may be carried out by the laboratory that generates the data and/or by the person or organisation that uses or interprets them. The most common errors in reports of analytical data are faults in transcription, such as incorrect positioning of the decimal point or transcription of data relating to the wrong sample. The number of times that data must be copied before the final report is composed should, therefore, be kept to a minimum. Other errors may include:

- Omission of a major ion. Analysis for nitrate, in particular, is sometimes omitted, although nitrate may represent a significant proportion of anions.

- Incorrect identification of samples.

- Double reporting of ions. This can occur, for example, when alkalinity titrations record ions that are also analysed separately, such as silicate and phosphate.

- Faulty analytical technique.

Some examples of basic data checks that may be readily employed are described below. Table 14.1 illustrates a series of validity checks on a set of results for some commonly measured variables in rivers.

Major ion balance error

Percentage balance error = 100 × (cations - anions)/(cations + anions).

For groundwaters, the error should be 5 per cent or less unless the total dissolved solids (TDS) value is less than 5 mg l^{-1}, in which case a higher error is acceptable. For surface waters, an error of up to 10 per cent is acceptable; if it exceeds 10 per cent, the analysis should be checked for errors in transcription or technique.

Table 14.1 Checking data validity and outliers in a river data set

Sample number	Water discharge (Q)	Elec. Cond. (µS cµ^{-1})	Ca^{2+} (µeq l^{-1})	µg^{2+} (µeq l^{-1})	Na^+ (µeq l^{-1})	K^+ (µeq l^{-1})	Cl^- (µeq l^{-1})	SO_4^{2-} (µeq l^{-1})	HCO_3^- (µeq l^{-1})	$\Sigma+^1$ (µeq l^{-1})	Σ^{-1} (µeq l^{-1})	NO_3-N (µg l^{-1})	PO_4-P (µg l^{-1})	pH
1	15	420	3,410	420	570	40	620	350	3,650	4,440	4,620	0.85	0.12	7.8
2	18	405	**3,329.4**	370	520	35	590	370	3,520	4,254	4,480	**0.567**	**0.188**	**7.72**
3	35	**280**	2,750	390	**980**	50	**1,050**	260	2,780	4,150	4,090	0.98	0.19	7.5
4	6	515	4,250	**5,200**	620	50	680	510	4,160	5,440	5,350	**0.05**	**0.00**	8.1
5	29	395	2,950	420	630	**280**	670	280	2,800	**4,280**	**3,770**	0.55	0.08	**9.2**
6	170	290	2,340	280	480	65	**930**	250	2,550	**3,165**	**3,730**	1.55	0.34	7.9
7	**2.5**	380	3,150	340	530	45	585	**3,240**	**375**	4,065	4,200	0.74	**3.2**	7.9

Questionable data are shown in bold

[1] Σ^+ and Σ^- sum of cations and anions respectively

Sample number

1 Correct analysis: correct ionic balance within 5 per cent, ionic proportions similar to proportions of median values of this data set. Ration Na^+/Cl^- close to 0.9 eq/eq etc.

2 Excessive significant figures for calcium, nitrate, phosphate and pH, particularly when compared to other analyses.

3. High values of Na^+ and Cl^-, although the ratio Na^+/Cl^- is correct - possible contamination of sample? Conductivity is not in the same proportion with regards to the ion sum as other values - most probably an analytical error or a switching of samples during the measurement and reporting

4 Magnesium is ten times higher than usual - the correct value is probably 520 µeq l^{-1} which fits the ionic balance well and gives the usual Ca^{2+}/Mg^{2+} ratio. Nitrate and phosphate are very low and this may be due to phytoplankton uptake, either in the water body (correct data) or in the sample itself due to lack of proper preservation and storage. A chlorophyll value, if available, could help solve this problem.

5 Potassium is much too high, either due to analytical error or reporting error, which causes a marked ionic imbalance. pH value is too high compared to other values unless high primary production is occurring, which could be checked by a chlorophyll measurement.

6 The chloride value is too high as indicated by the $Na^{+/}Cl^-$ ratio of 0.51 - this results in an ionic imbalance.

7 Reporting of $SO4^{2-}$ and HCO_3- has been transposed. The overall water mineralisation does not fit the general variation with water discharge and this should be questioned. Very high phosphate may result from contamination of storage bottle by detergent.

An accurate ion balance does not necessarily mean that the analysis is correct. There may be more than one error and these may cancel each other out. As a result additional checks are needed.

Comparison of electrical conductivity and total dissolved solids

The numerical value of electrical conductivity (µS) should not exceed that of TDS ($mg\ l^{-1}$). It is recommended that conductivity be plotted against TDS and values lying away from the main group of data checked for errors. The relationship between the two variables is often described by a constant (commonly between 1.2 and 1.8 for freshwaters) that varies according to chemical composition. For freshwater the normal range can be calculated from the following relationship:

Conductivity ≈ TDS × *a* where *a* is in the range 1.2-1.8

Typically, the constant is high for chloride-rich waters and low for sulphate-rich waters.

Total dissolved solids (calculated) v. total dissolved solids (dry residue)

The calculated value of TDS and that of TDS determined by dry residue should be within 20 per cent of each other after correction for loss of CO_2 on drying (1 g of HCO_3 becomes approximately 0.5 g of CO_2). However, dry residue can be difficult to determine, especially for waters high in sulphate or chloride.

It must be scientifically possible

Totals of any variable must be greater than the component parts as in the following example:

- Total coliforms must be greater than thermotolerant coliforms.

- Total iron must be greater than dissolved iron.

- Total phosphate must be greater than dissolved phosphate.

- Total phosphate must be greater than phosphate alkalinity.

The species reported should be correct with regard to the original pH of the sample. Carbonate species will normally be almost all HCO_3^-, high CO_3^{2-} levels cannot exist at a neutral pH and CO_3^{2-} ions and H_2CO_3 cannot coexist.

Departure from expected values

Distinguishing between a valid result and an error in the data requires experience. Visually scanning a spreadsheet of data can help to identify analyses in which some variables are much higher or lower than in others, although the values may not necessarily represent errors and may be of interest in their own right. For analytical techniques it is recommended that control charts (see Chapter 9) be prepared to identify values than lie away from the main group of data. Otherwise, specialised techniques for identifying outliers need to be used (see *Water Quality Assessments* for details).

Anomalous results

If nitrate is present in the absence of dissolved oxygen, the value for one or the other is likely to be incorrect, since nitrate is rapidly reduced in the absence of oxygen. There may have been a malfunction of the dissolved oxygen meter or loss of oxygen from the sample before analysis.

If nitrite and Fe_2^+ are present in the same sample, either the analysis is incorrect or the sample is a mixture of aerobic and anaerobic waters. For a lake, it could be a representative sample, but for a groundwater it would probably mean that the water was derived from two different aquifers.

If there are high levels of iron or manganese and of aluminium (of the order of 1 mg l^{-1}) at a neutral pH, they are probably the result of colloidal matter which can pass through a 0.45 μm filter.

Analytical feasibility

If the results do not fall within the analytical range of the method employed, errors of transcription or of analytical technique should be suspected.

14.2 Data handling and management

The computer software used in data handling and management falls into three principal classes:

Statistical software which processes numerical data and performs statistical tests and analyses.

- Spreadsheets which handle both numerical data and text, and usually include powerful graphical and statistical capabilities (thus overlapping with the purely statistical software).

- Database software which is designed to manage the input, editing and retrieval of numerical data and text. No statistical or graphical capabilities are in-built, but the power of the programming language allows the skilled user enormous scope for data manipulation, sorting and display.

In an ideal situation all three classes of software can be used together in a complementary fashion. Nevertheless, it is essential that the eventual statistical methods to be applied are taken into account fully in the original monitoring programme design. This should help to ensure that the data produced are adequate for the statistical techniques which are to be applied.

A fourth class of data-handling software, known as GIS (geographical information systems), has recently been developed. This is specifically designed to relate data to geographical locations and output them in the form of maps. Data on different aspects of, in this case, a water quality monitoring programme can be superimposed. For example, data on groundwater quality such as chemical variables can be overlaid with data on land use, etc. This allows the relationships between the selected aspects to be studied and a map, for example of groundwater protection zones, to be generated. The overwhelming constraints on the wider use of GIS software are its cost and its need for sophisticated hardware and highly skilled operators.

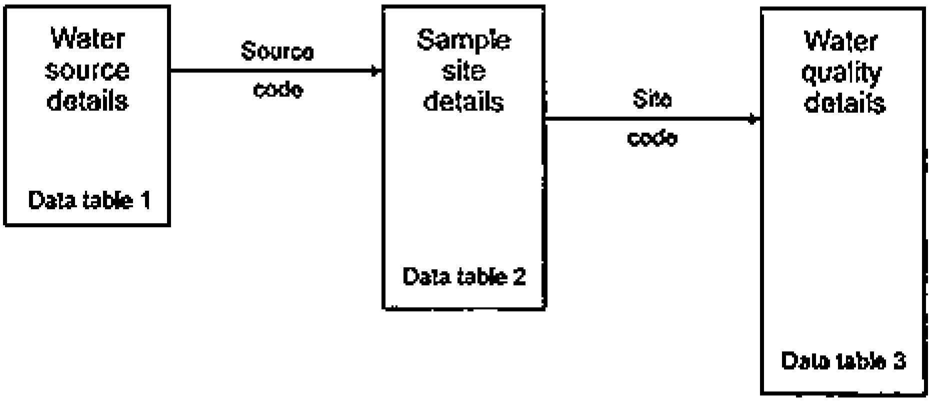
The scope and nature of computerised data-handling processes will be dictated by the objectives of the water quality monitoring programme. Generally speaking, however, a database offers the best means of handling large quantities of data and it should be capable of exporting data in formats that are accepted by all good statistical, spreadsheet and GIS packages.

An experienced programmer can create a relational database management system (RDBMS). Such systems allow the creation of data tables that can be related to other, associated, data tables through unique keys or indexes. A data table is a collection of data records and a record comprises a series of data fields or information variables. A simple analogy for a database is provided by a telephone directory in which each record, or entry, consists of various information fields, i.e. family name, initials, address, etc., and telephone number. Entries are listed in alphabetical order of family name and then of initials (or forenames). However, since many people may share the same family name, the entry “Smith”, for example, cannot be considered a unique index. The only unique index or key in the directory is the individual telephone number.

In creating a database, the first decision to be made concerns what fields of information are required and the second, how many data tables are needed for their efficient storage. It is a mistake to group all data in a single table unless, like the telephone directory, the database is to be very simple. A useful approach is to separate fields of information that relate to variable data and those that deal with unchanging details, such as the name, description and location of a sampling site.

All data tables relating to a single sampling site must be identifiable as such, regardless of the type of data they contain. To avoid the need for including extensive details of site description and location in every such table, each sampling site should be identified by a unique index, usually a code of no more than 10 characters, which appears in every data table relating to that particular site. Then, if it becomes necessary to amend any details of site description, only one data table, the one concerned with the site itself, will need to be changed. All other related tables will continue to carry the unique, and unaltered, code.

Figure 14.1 Relationship between tables for a simple water quality database



Other unique indexes, or keys, can be created to relate data tables to each other. It may be valuable to create a key based on sampling dates, for example, which would allow rapid retrieval of all data recorded for a particular date or between two specified dates. Consequently, it becomes obvious that the desired format of the output from the database dictates the keys that should be created, such that selecting the appropriate key or index yields data in a particular order or for a particular variable.

Figure 14.1 illustrates the linking of tables described above by providing a suggested scheme for a simple water quality database. Three data tables have been used to store the required information. The first table is concerned with the water source to be tested, which might be a lake, river, borehole, water supply scheme or even an industrial treatment plant. The name, type and location of the water source are recorded in this table, and possibly a brief description. In addition, the table must contain a source code, the primary “look-up” key, that links it to the second table.

The second table contains more detailed information about sampling sites. A river, for example, may have two or more designated sampling sites, the locations for which must be accurately defined. This table carries the source code or look-up key from table one, plus unique codes for each sampling site that provide links to the third table. Typically, the second table would contain the following information for each site:

- Source code: the look-up key to table one.

- Site code: the link to the third table.

- Location: geographical location of the sampling site.

- Map number: a reference to the map covering the site location.

- Map reference: the latitude and longitude of the site or a national grid reference.

- Description: a brief textual description of the site.

Data on water quality (temperature, pH, conductivity, chemical and microbiological variables, etc.) are stored in the third table. Many details will be recorded for each sample from each site. Additional information should cover how and when each sample was collected and when the various analyses were performed. Ideally, each sample record should include a unique key for purposes of identification, cross-reference, and data retrieval; for example, this might be the analysing laboratory’s reference number for the sample.

Certain precautions relating to the deletion and/or amendment of information must be rigorously observed. In the example of Figure 14.1 deletion of a record from table one might leave related data in table two “orphaned”, that is, without linkage to “parent” data. Deletions should, therefore, be made only of material which is not shared by a related table. Moreover, amendment of a key field in one table must be reflected in parallel changes to related records in other tables. These simple rules are essential for preserving the referential integrity and validity of the database, both of which are crucial to efficient data management. Many modern databases track and maintain their referential integrity, thereby reducing the likelihood of creating orphaned data tables.

A final point concerns the designs of screens for the input of data to the database. Where standardised forms already exist for written records, these should be mimicked as far as possible for computerised data entry. The greatest advantages of the computerised database are ease of manipulation, ease of retrieval and, particularly, ease of dissemination of data to all interested parties. It is this last feature, which is one of the principal objectives of the water quality monitoring programme, that gives computerised databases their enormous advantage.

14.3 Basic statistical analysis

Producing meaningful information from raw data normally requires an initial statistical analysis of that data; for example, to determine the magnitudes of variables, their variability, any time trends, etc. It is also necessary to be able to give some indication of the "confidence" the user may have in the statistical outputs. For example, if a sample average is 50 per cent greater than a previous estimate, could that be expected as a reasonable chance occurrence, or is it indicative of real change? In order that such results are valid, it is essential that only statistical tests which are suited to the data (defined by its type, manner and frequency of collection, etc.) are used. With the ready availability of computerised statistical packages it can be too easy to apply inappropriate analysis to data, especially when they can be automatically collected, stored and presented for analysis. This handbook does not aim to provide detailed methods for statistical analysis, but only to show the potential of such analysis and of effective presentation of data. Worked examples of some of the statistical techniques discussed here are available in *Water Quality Assessments*.

14.3.1 General considerations

Many powerful, traditional statistical tests rely on the data conforming to an underlying pattern or frequency distribution. Most commonly, this is the "normal" distribution, or variants thereof. Such statistical tests are termed parametric to indicate the requirement that the data conform to some understood and describable (by its parameters) distribution. The mean and standard deviation, for example, are of this type. Another fundamental class of statistical tests are non-parametric (or distribution-free) which, as their description implies, make no assumptions about the data to which they may be applied. The median and the percentiles are examples of non-parametric tests.

When data conform to the appropriate distribution requirements, the parametric statistical tests are usually more powerful than their non-parametric counterparts (where they exist). However, any advantage can be swiftly eroded once the data distribution becomes distorted and cannot be corrected by transformations, for example by using logarithms of the data instead of the raw data. Water quality data sets are often not easily definable, therefore there has been much recent development of non-parametric statistics, both to provide complementary tests to their parametric counterparts and to improve their power. An important, early step in programme design is, therefore, to assess which statistical tests would best (and most simply) serve the information needs of the programme, to evaluate any data requirements this may imply and, thus, to ensure that the monitoring programme is designed to provide for those data needs.

14.3.2 Outline of statistical techniques

Statistical analysis of water quality data may be very broadly classified in two groups:

- Descriptive statistics.
- Inferential statistics.

These classifications suggest that, initially, the task is to provide an accurate and reliable statistical description of the data set. Following that it may be possible to give some account of how well the data set, and/or its interactions, conforms to some theory as to its origin and so, possibly, identify potential means of modification. Thus the "what?" comes before the "how?".

The basic statistical techniques which constitute these two classes can be generally identified as:

- Analysis of data distributions.
- Testing assumptions about data sets.
- Specifying data magnitudes and variability.
- Estimating reliability of data statistics.
- Comparisons of data sets.
- Associations between data sets.
- Identifying trends and seasonality within data sets.
- Testing theories relating to the water quality data.

The various tests could be further sub-divided by virtue of being parametric or non-parametric, single or multi-variable, single or multi-sample, discrete or continuous data, etc.

Analysis of data distributions

Analysis of data distributions is normally carried out to check that the data set, either in its raw state or following transformation, conforms to a definable data distribution which is appropriate to the parametric statistics to be applied to it. Knowing the distribution may help in accepting or rejecting assumptions about the sampling conditions which generated the data and may also allow some predictions to be made about the data. Formal analysis of distribution can be a difficult task, particularly with sparse and variable data. However, the first activity (as with almost all water quality data) is to plot the data in an appropriate manner (e.g. scatter plot, frequency histograms, probability plots, see section 14.5). This will often give a reasonable indication of the distribution type, even though statistical difficulties may be encountered. Usually, once a general idea of the distribution is gained, the assumption would be tested against the data available.

Testing assumptions about data sets

The tests in this category usually involve procedures which assess the statistics derived from some other test against values which are characteristic of some presumption about the data. For example, are the frequency histogram data consistent with a normal distribution? Traditionally the c^2-test has been used, but other tests (such as the *W*-test, see *Water Quality Assessments*) provide some advantages over this traditional approach, depending on the nature of the data.

Specifying data magnitudes and variability

This category would include all the analyses leading to the mean, variance, standard deviation, and coefficient of variation (parametric) statistics, together with the median and appropriate percentiles of non-parametric statistics. The 25th and 75th percentiles are the most commonly used. The 25th and 75th percentiles are the values below which 25 per cent and 75 per cent, respectively, of the data values occur.

Estimating reliability of data statistics

The purpose of these methods is to provide some measure of the possible errors which may be associated with the estimated statistics. Thus, for example, the standard error of a mean allows some judgement of how great a difference in mean may occur in a repeated sampling. The availability of such error estimates allows confidence limits to be constructed for a variety of statistics, and from which some conclusions about the reliability of the data may be drawn. Other procedures in this category would be to deal with outliers. These are data which do not conform to the general pattern of a data set. They may be anomalous, in which case they must be excluded because of the distortion (particularly for parametric tests) they may cause. However, they may also be highly informative, proper, extreme values of the data set and excluding them would lead to incorrect conclusions. Critical testing is, therefore, necessary for such data.

Comparisons of data sets

Monitoring programmes often attempt to compare water quality conditions at one site with that at another (perhaps acting as a control site). Various statistical methods are available to compare two or more such data sets. These range from such basic tests as the Student's *t*-test to the Mann-Whitney *U*-test and the Kruskal-Wallis test for multiple comparisons.

Associations between data sets

Once a programme begins to address the issues of forecasting, or establishing causes-and-effects in water quality variables, then it is essential to be able to quantify the relationships amongst the variables of interest. Two primary classes of such assessment are correlation analysis and regression. Correlation establishes the statistical linkage between variables, without having to ascribe causation amongst them whereas regression seeks the functional relationship between a supposed dependent water quality variable and one, or many, variables which are termed "independent".

Identifying trends and seasonality within data sets

Water quality variables frequently exhibit variability in time. This variability may be cyclical with the seasons, or some other established variation over time. Special statistical tests have been developed to deal with these possibilities. Many are complex and too advanced for a basic water quality monitoring programme. Nevertheless, the Mann-Kendall test for trends, and the Seasonal Kendall test for seasonality, are amongst the more general tests which may be applied.

Testing theories relating to the quality data

Once basic data have been analysed and tentative conclusions have been drawn, almost all water quality monitoring programmes require that those conclusions are validated as far as possible. Generally, the approach is to define null-hypotheses, broadly proposing the alternative to the conclusion of interest. The null-hypothesis is then subjected to testing using the statistics available. If refuted, then the hypothesis of interest is accepted within the probability defined for the testing procedure. The test procedures available include the *t*-test, *F*-test and the analysis of variance (ANOVA).

This brief review has allowed only very few, basic methods to be identified. Nonetheless, many can routinely form part of sensitive statistical analyses of water quality data. Almost all methods mentioned will be available as part of computer statistics packages, in addition to their detailed description in applied statistical texts (see section 14.8).

14.4 Use of data and the need for supporting information

The achievement of the common objectives of monitoring programmes cannot rely solely on water quality data and the derived statistics. Information on water quality must be placed in the context of the natural and man-made environment and may need to be combined with quite different types of data. This is particularly true if the ultimate intention is regulation or control of water quality. For example, a monitoring programme might be designed to assess the effectiveness of pollution control of domestic sewage in relation to the quality of well waters. It is relatively easy to visit a number of wells, take samples, produce laboratory data and report the degree of contamination of the different samples. However, this does not constitute assessment; it merely provides some of the information needed for assessment. Assessment of pollution control would also require information on the different strategies for sewage collection and treatment, and for source protection at the different sites. It would then be necessary to determine whether the patterns of waste generation were similar and whether the groundwater bodies were comparable in their hydrogeology. While it might seem relatively easy to identify deterioration or improvement at a particular site, it is far more difficult to explain exactly why those changes have occurred and to suggest actions for management. The changes that followed management action would also need to be monitored in an effort to identify the most effective action.

The above example is intended to show that water quality data must often be combined with other data and interpreted in a way that specifically addresses the objectives of the end-user of the information. Some water quality monitoring programmes generate excellent data but lack supporting information on the possible influences on water quality. This poses a problem for management because control of water quality frequently requires intervention in human activities, such as agriculture, industry or waste disposal. Point sources of pollution may be easy to identify, but effective control of diffuse sources requires more extensive data. If the information about the influences on water quality at a particular location is incomplete, it may not be obvious where management action would be most effectively deployed. Water quality assessment and management is, therefore, concerned not only with monitoring data but also with a wide range of supporting and interpretative data that may not always be available. Sometimes priorities for action must be selected on the basis of limited information.

Water quality standards and indices

Pure analytical data may not always convey to non-experts the significance of poor water quality or the possible impacts of elevated chemical concentrations. The typical administrator may have a limited technical background, and it may be necessary to use special “tools” to communicate relevant information in a manner that will prompt suitable interventions.

The simplest tool is a chemical standard or “trigger” value. When the trigger value is exceeded, the administrator (or other information user) is prompted to take action. That action, for example, may be designed to prevent people from drinking water that does not meet health standards or it may be a management intervention to regulate discharges in an operational situation. A variant of the trigger approach is based on a percentage of values exceeding the standard. When a certain number of values (e.g. 5 per cent) fall outside an acceptable standard, the water is judged to be of unacceptable quality for specific uses.

Standards are also important as objectives for quality. They may function as targets for quality control programmes, and the relative success of a management strategy may be judged by comparing the monitoring data with established standards. Standards are, therefore, convenient means of stating expert judgements on whether water is of an acceptable quality for a particular use.

Information may also be summarised and presented to the non-expert in the form of water quality indices. A water quality index combines data for certain variables that have an impact on water quality. An index may simply aggregate the values or it may score or weight data according to their relative importance for water quality. Examples include general water quality indices (combining a range of chemical and physical variables), use-oriented indices (including variables of importance for specific uses) and biological indices (combining scored information for various species sensitive to changes in water quality, see section 11.2).

14.5 Simple graphical presentation of results

Graphs can communicate complex ideas with clarity, precision and efficiency. They will reveal patterns in sets of data and are often more illustrative than statistical computations. The general objective in graph construction is to concentrate a large amount of quantitative information in a small space so that a comprehensive overview of that information is readily available to the viewer. This can be achieved if:

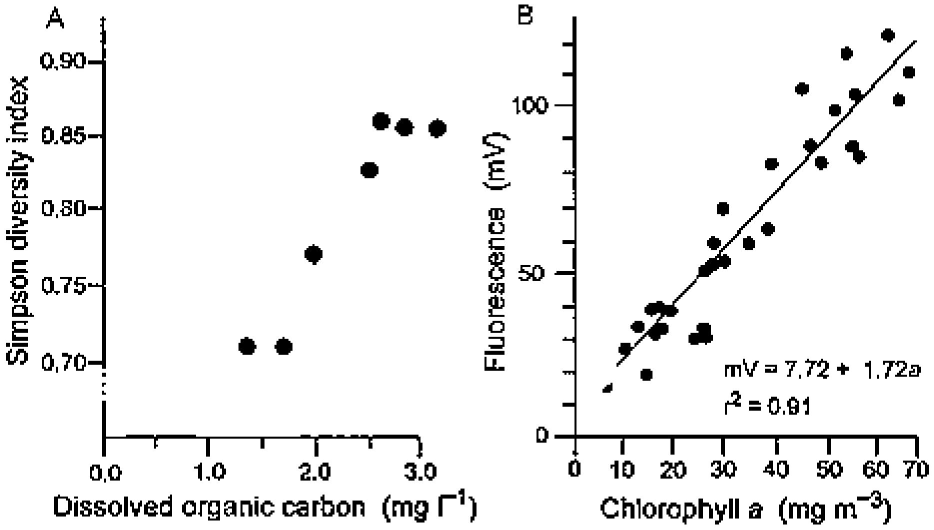
- the graph is uncluttered,

- lines, curves and symbols are clear and easy to see,

- appropriate scales are chosen so that comparison with other graphs is possible, and

- clearly different symbols are used to represent different variables.

Figure 14.2 Example of scatter plot. A. Correlation between Simpson diversity Index and dissolved organic carbon (After Pinder, 1989) B. Regression of chlorophyll *a* and fluorescence (After Friedrich and Viehweg, 1987)



There are a wide range of graphical methods which are suitable for reporting data from monitoring programmes. In deciding the most appropriate form of graphical presentation, four basic types of plot are usually involved: a scatter plot, a bar graph (histogram plot), a time series plot, and a spatial plot. All of these may be presented in several different ways. Time series plots are designed to illustrate trends with respect to time, together with any seasonality effects. Spatial plots can be used to illustrate vertical and longitudinal profiles and, by using maps and cross-sections, can demonstrate geographic and local quality distributions. Survey results are often very effectively displayed on maps. These basic graph types can be used with both raw data and derived statistics. There are also other plot types, such as box plots, pie diagrams and trilinear diagrams (rosette diagrams) which are most suited to displaying summary data. It may also be necessary, in some instances, to display supporting, ancillary information such as river flows, sample numbers and frequency, etc.

Examples of scatter plots are shown in Figure 14.2. They each illustrate two quality variables sampled at the same time. There may, or may not, be implied cause-and-effect in such graphs. At its most basic, the scatter plot gives an indication of the co-variation (or lack-of) between two variables. The scatter plot matrix is a development of this basic concept that can be applied to a variety of different data types and to illustrate differing responses. It is a simple, but elegant, solution to the difficult problem of comparing a large collection of graphs. In the scatter plot matrix, the individual graphs are arranged in a matrix with common scales. An example is shown in Figure 14.3. A row or column can be visually scanned and the values for one variable can be seen relative to all others. Thus column A shows the variation of conductivity, TON, SiO_2 and SO_4 with flow, whereas row B shows the conductivity associated with flow, TON, SiO_2 and SO_4. Such a graph can often demonstrate the significance of the data much more dramatically than extensive statistical analyses.

A bar graph is shown in Figure 14.4. The example deals with summary values, and so confidence limits may also be applied to the top of the bars to give an indication of the variability of the data. In another type of bar graph the bars may also be “stacked” in order to display relative proportions.

Figure 14.3 A scatter plot matrix for five variables

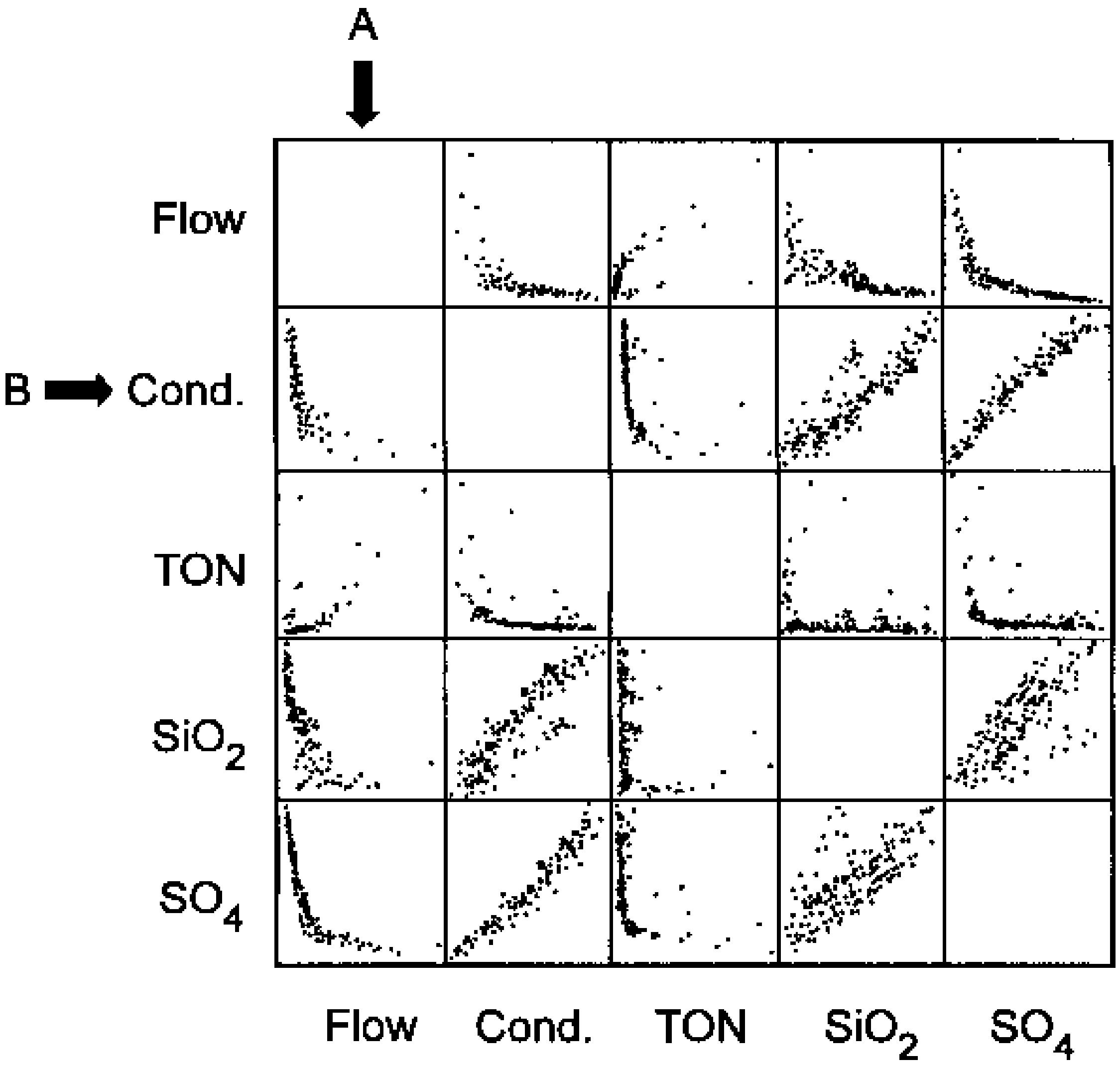


Figure 14.4 Example of a bar graph with 95 per cent confidence limits indicated as vertical lines (After Spodniewska, 1974)

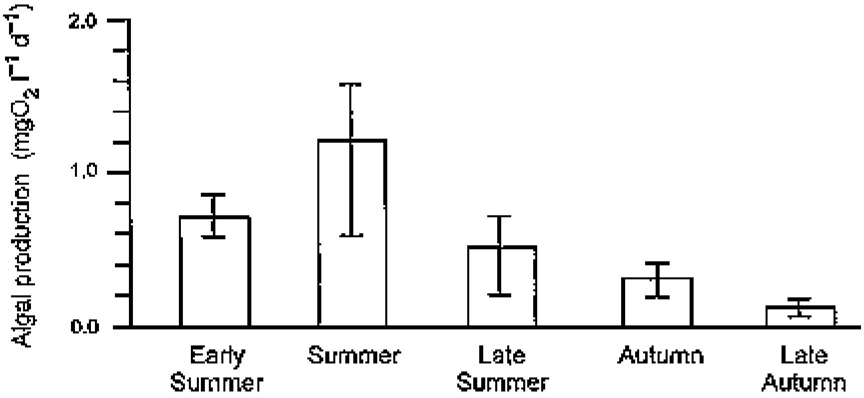


Figure 14.5 A time series plot showing a long-term trend (After Humpesch, 1992)

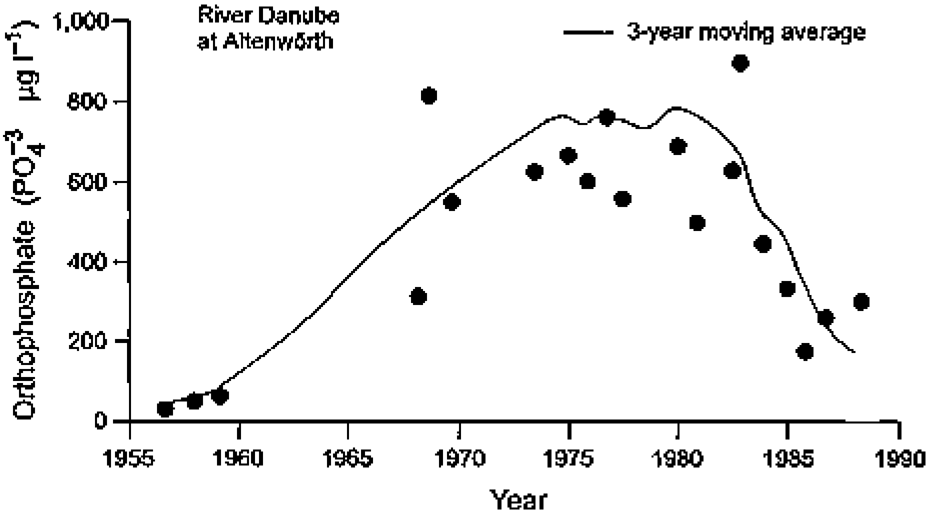
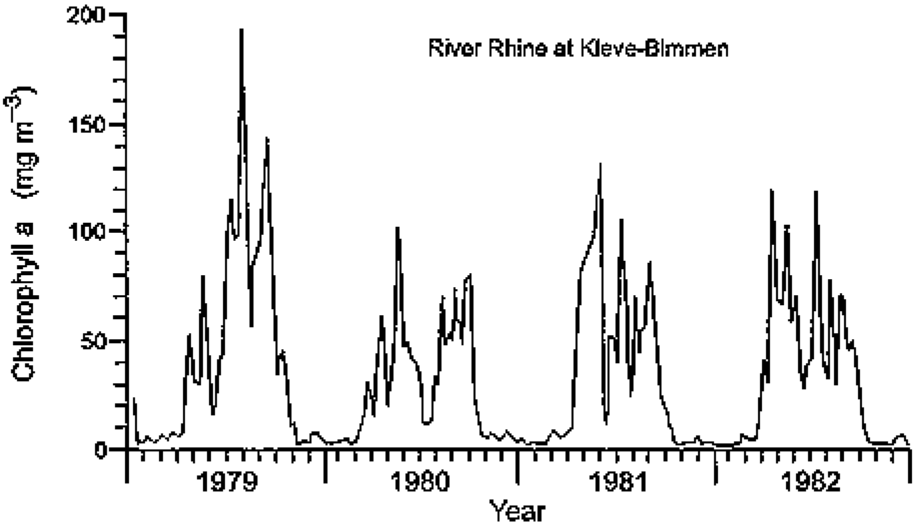


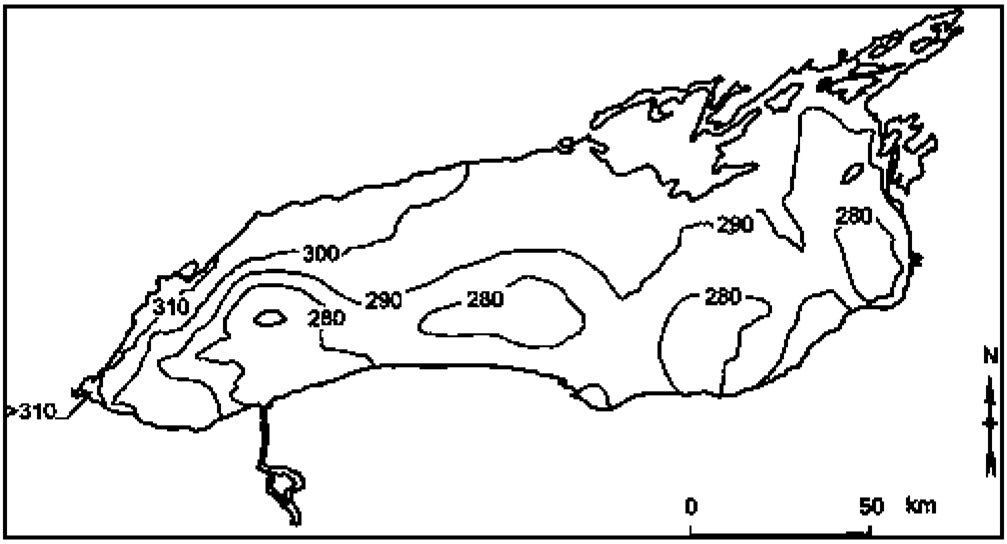
Figure 14.6 A time seriesplot indicating seasonal variations (After Friedrich and Viehweg, 1984)



Figures 14.5 and 14.6 show different uses of time series plots. The concentrations of the variables in a series of samples collected over a time period are plotted against the times at which the individual samples were obtained. These show clearly how the concentrations of the variables change with time and can be used to indicate possible trends (Figure 14.5) and seasonal variations (Figure 14.6). The information which may be gained is governed by the frequency of sampling in relation to the variability of the water quality characteristics being studied.

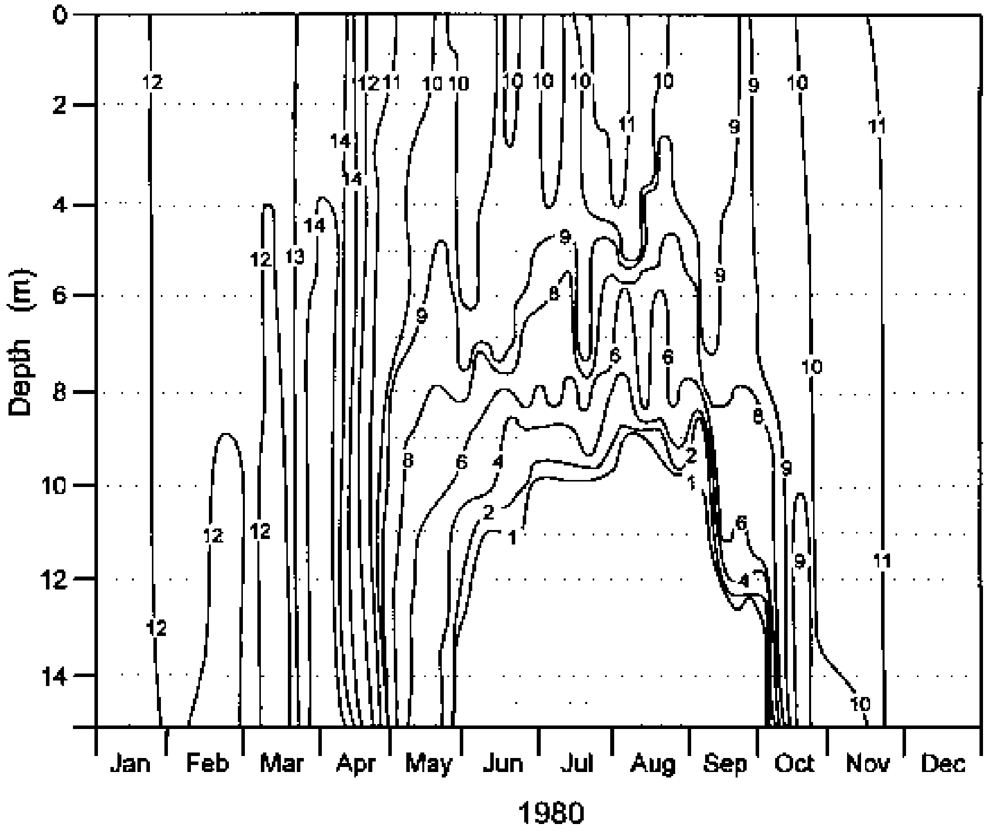
Spatial plots can be used to show horizontal or vertical variations in water quality variables. Figure 14.7 shows an example of the horizontal distribution of conductivity in a large lake.

Figure 14.7 A spatial plot showing isopleths of conductivity in Lake Ontario



Profiles through space or time summarise large quantities of information in a single figure. Interpolation between data points to obtain isopleths (lines joining points where the value of a variable is the same) as in Figure 14.7 may be done either manually or by computer. This type of graph can also combine temporal and spatial data, and is particularly useful, for example, for illustrating seasonal changes in the vertical composition and characteristics of lake waters (Figure 14.8).

Figure 14.8 Combination of spatial and temporal data to show vertical variations in a lake. Isopleths of oxygen in a lake which is seasonally stratified



The findings of a survey can be effectively displayed by superimposing the monitoring results on a map of the area surveyed. The survey results in each sub-area may then be presented in a graphical form (framed rectangle, pie chart, etc.), in which the measured value of the variable is in proportion to some identifiable value. In the example of framed rectangle graphs in Figure 4.9, the population dependent on groundwater is shown as a percentage of the total population in each sub-area. This example also shows the dependence on groundwater for the entire area, so that the relative dependence in each administrative subdivision is immediately obvious. If a water quality variable was the item of interest, its measured concentration could be shown in relation to the water quality standard or some other absolute value for that variable. Some reference must always be provided so that the viewer can readily understand the significance of the reported result.

Figure 14.9 Example of framed rectangle graphs on a map

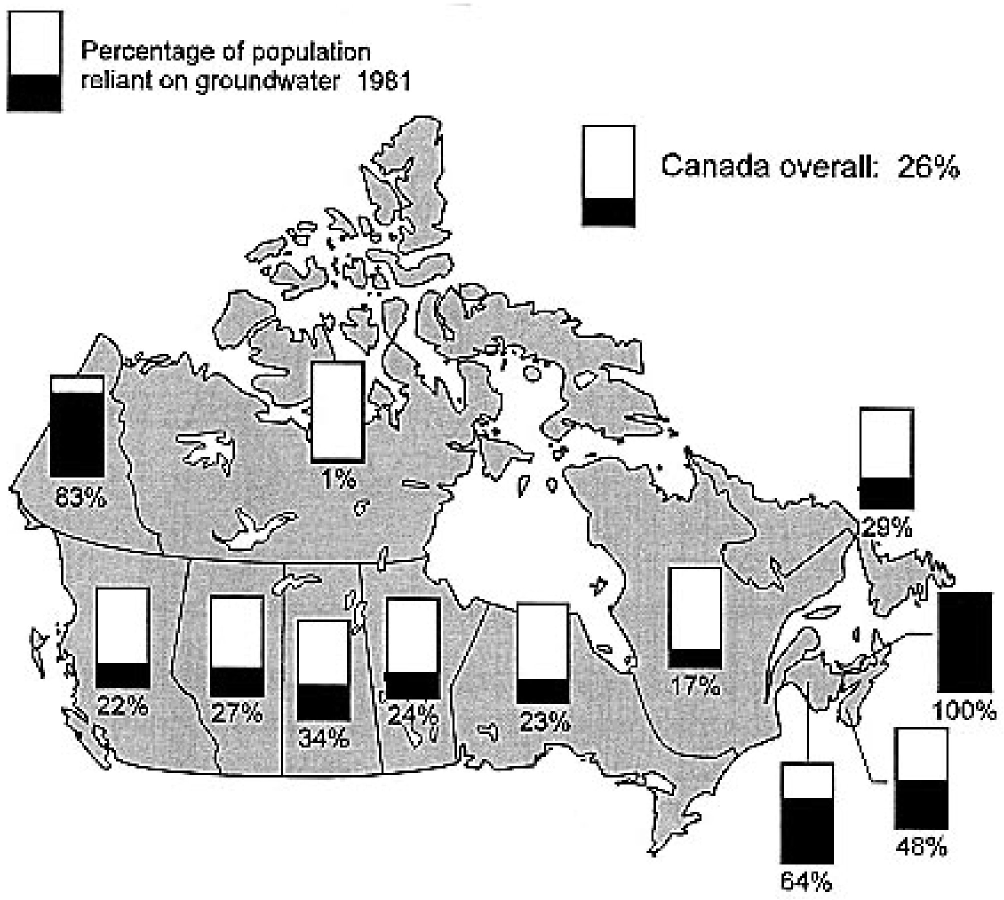
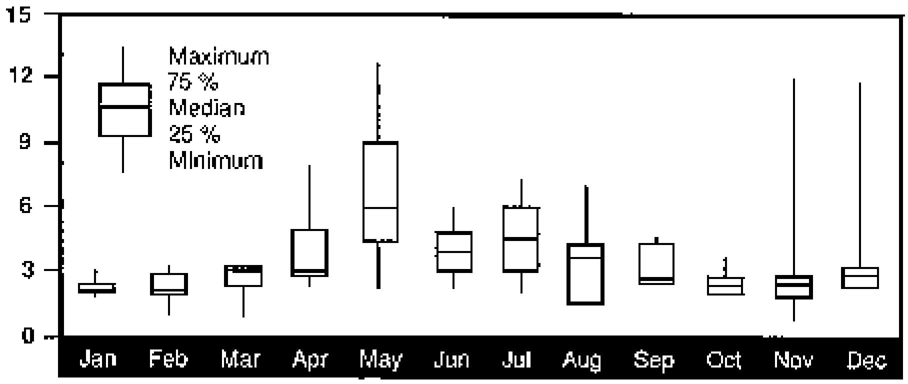


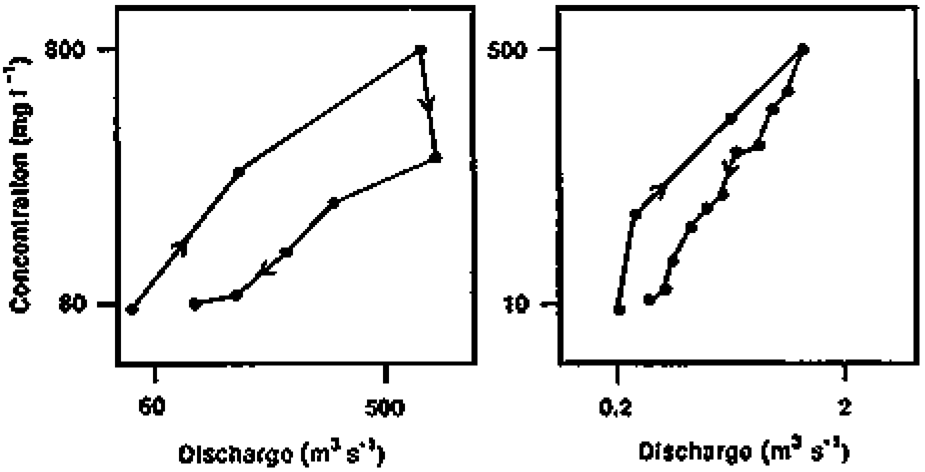
Figure 14.10 The use of box and whisker plots to display time series information. Inset shows the construction of a box and whisker plot



Box and whisker plots are a useful way to compare distributions because they permit the comparison of corresponding percentiles. They are used to indicate the basic statistical attributes (maximum, minimum, median, upper and lower quartiles) as shown in the inset to Figure 14.10, which illustrates the use of such plots to display time series information.

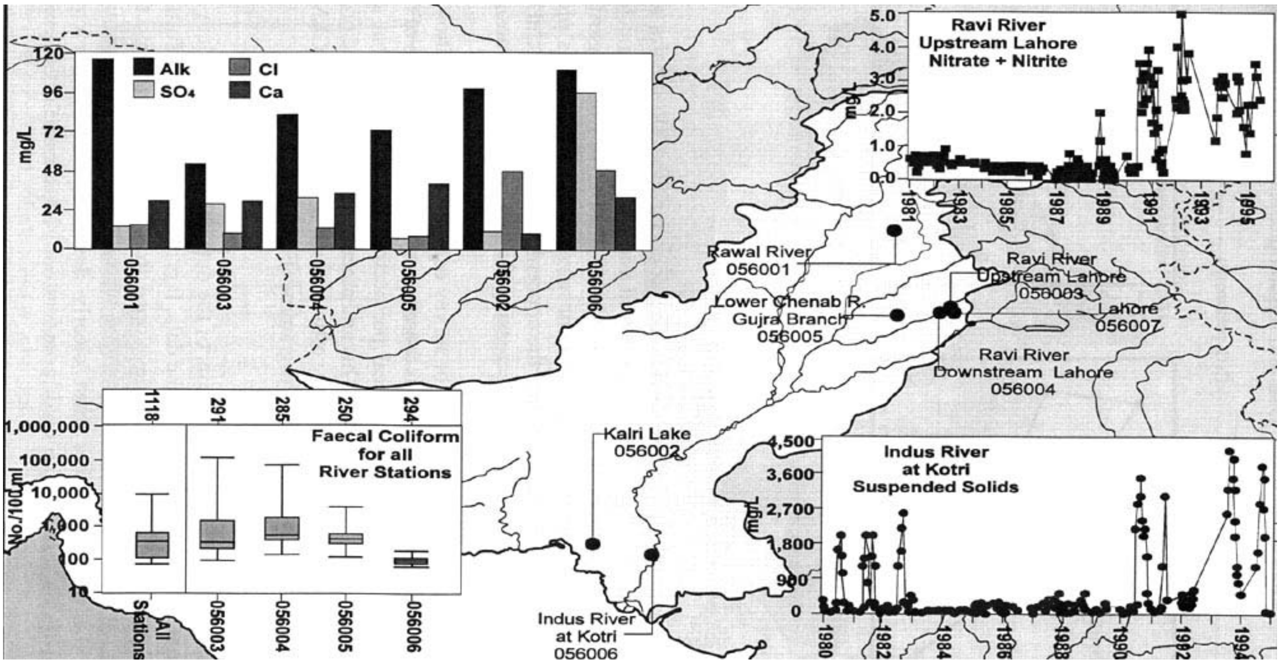
The discharge volume of a river is a major factor in understanding the river system. In many rivers, flow will change with time by as much as several orders of magnitude. If discharge is plotted on a time series graph, flood peaks will dominate the discharge scale but will be infrequent on the time scale. Similar features may exist for the concentration of one or more variables in a river, such as suspended sediment concentration. The relationship between concentration of a variable and river discharge is clearly demonstrated in a hysteresis plot in which concentration is plotted against flow as shown in the example of Figure 14.11. The concentration-flow pairs are plotted in the order in which they are observed and the points are connected in the same order. In many cases the graph is more illustrative when logarithmic scales are used for both flow and concentration. The hysteresis plot, showing each data point, indicates that the concentration of a variable can be different when stream discharge is decreasing from what it is at times of increasing discharge.

Figure 14.11 Examples of hysteriasis plots of concentration versus discharge



A very useful overview technique for survey data is the superimposition of other types of graphical analyses (such as those presented above) on a map of the study area. This permits a geographical perspective on differences that are observed at different monitoring sites. This and other types of analysis and presentation are easily done with computer software such as the RAISON/GEMS package used in the GEMS/WATER Programme (Figure 14.12).

Figure 14.12 Presentation of spatial and time series data on a map of the study area. Water quality indicators for the Indus and Ravi rivers in Pakistan. Alkalinity is reported as mg l^{-1} $CACO_3$. Nitrate is reported as mg l^{-1} NO_3^-. Data assessment by the UNEP GEMS Collaborating Centre for Freshwater Monitoring and Assessment, Canada. Analysis was undertaken using the GEMS/RAISON system.



14.6 Reporting

Reporting should focus on the synthesis of the data collected, rather than on the individual numbers that make up the data. It is important that the resultant interpretation gives the broad view developed from, and supported by, the fine details. In order to convey information effectively to managers, politicians and the public, reports must clearly describe the environmental situation in terms that can be readily understood. Most people who read reports on monitoring are not specialists and simple, clear explanations are essential for effective communication.

Monitoring results, reported in terms of the effects that processes, events and mechanisms have on water quality, should be readily understandable to their diverse audiences. Authors must recognise the broad range of readership and endeavour to present material in the simplest possible terms.

14.6.1 Types of report

The several different types of report that are possible have certain elements in common, including the statement of objectives and the description of the study area. To avoid any misunderstandings it is important that all elements of the monitoring programme are precisely and clearly described. The four principal types of report are described below.

Study plan report

The study plan report (or monitoring programme document, see Chapter 3) defines the objectives of the monitoring programme, including the questions to be addressed and the present understanding of the environment to be studied. It also defines the sampling and data review strategy that will be followed to meet the objectives. In this regard it must be noted that data collected for a specific purpose may not be applicable to other issues and data collected for no clearly stated purpose may not answer any questions at all. If scarce resources are to be used effectively, it is essential that a study plan is prepared. The study plan is analogous to the workplan for a building, i.e. it clearly documents the processes so that all parties understand their roles and adopt common goals for the overall study.

Protocol and methods report

The protocol and methods report, sometimes known as the Standard Operating Procedure (see Chapter 9), describes methods and equipment in sufficient detail for other scientists to be able to assess the scientific validity of the results reported. Without this documentation, work might be viewed as merely semi-quantitative and the conclusions regarded as suspect, if not completely invalid. Detailed descriptions of all methods must be included, i.e. sample collection, sample preservation in the field, field analyses, laboratory procedures, quality assurance procedures, data recording, and computer processing. All reporting procedures should be covered, including those for reporting quality assurance results and the results of various analytical measurements. The report should include provisions for the protection of all data against loss or damage by the regular filing of duplicate copies in a safe and secure location.

The report should be prepared at the beginning of a programme and must be regularly reviewed, revised and updated. Review helps to ensure the detection of problems that could invalidate the results. International protocols should be used to the greatest possible extent. All members of the study team should contribute to the preparation of the protocol and its subsequent revisions because collective knowledge and understanding of procedures enhance the quality of the data collected.

Data report

In some circumstances it may be necessary to report data without detailed interpretation, and a data report permits the early distribution of information before interpretation is complete. However, it may contain unverified or uncertain results and, as a result, will become outdated as the main body of data is corrected and amended. If data are stored on a computer, data reports can distributed in machine-readable form.

The primary purpose of a data report (or indeed of any report) is to transmit information to an audience. The information should be assembled in a well organised format so that the reader can easily review it. This is particularly true of a data report, which will usually consist of tabulated material with few supporting statements. Suppose, for example, that samples taken at a site once a month for a year were analysed for 30 variables and that the results formed the basis of a data report. If this report were arranged with all 30 results for each sampling date on a separate page, a reader looking for seasonal trends would have to turn pages 12 times per variable. If, however, the results were arranged chronologically in columns, patterns or extreme conditions could be readily observed. Graphical presentation and statistical summaries can also be used to improve understanding of data.

Interpretative report

Interpretative reports provide a synthesis of the data and recommend future actions. Different approaches may be used, for example process identification, problem-solving, and achievement of desired goals, are some alternatives. The synthesis of the data should be targeted to the specific objectives for which the data were collected. Interpretative reports should be produced regularly to ensure that programme objectives are being met and are currently valid. The time interval between successive reports should never be greater than two years.

14.6.2 Structuring a report

The structure described in the following paragraphs may be used as a general outline for the preparation of monitoring reports, although not all reports will necessarily contain all of the elements.

- *Summary*: The summary briefly outlines what was done, describes the significance of key findings, and lists the important recommendations. It should be written so that a non-scientist can understand the purpose of the work and the importance of the results. Its length should never exceed two or three pages (about 1,000 words). Often, the summary is the only part of the report which is read by senior managers.

- *Introduction*: The objectives and the terms of reference of the study should be presented in the introduction. The problems of issues being addressed should be clearly stated, previous studies should be described, and relevant scientific literature should be reviewed. Any major restrictions (personnel, access, finance, facilities, etc.) which limit the desirable scope of the programme should be identified.

- *Study area*: A summary of the geography, hydrology and other salient features (which, for example, may include land use, industry, and population distribution) of the area under study should be provided (see section 3.5). All locations, structures and features mentioned in the text should be identified on maps such as the example shown in Figure 14.13. In certain cases, profiles of river or lake systems may provide useful information to augment that provided on the maps.

- *Methods*: A separate section should provide details of the methods and procedures used for all aspects of the study. In most cases this can be covered by reference to the protocol and methods report (see section 14.6.1 and Chapter 9). A summary of the procedures to be followed for quality assurance should be presented (see Chapter 9).

- *Results*: Results should be presented in graphical form whenever possible, and graphs should clearly demonstrate the relationship of the data to the monitoring objectives. Some examples of suitable types of graph are shown in section 14.5. Graphs should summarise entire sets of data rather than individual observations. International units of measurement should be used wherever possible but if local units are used precise descriptions and conversions should be provided.

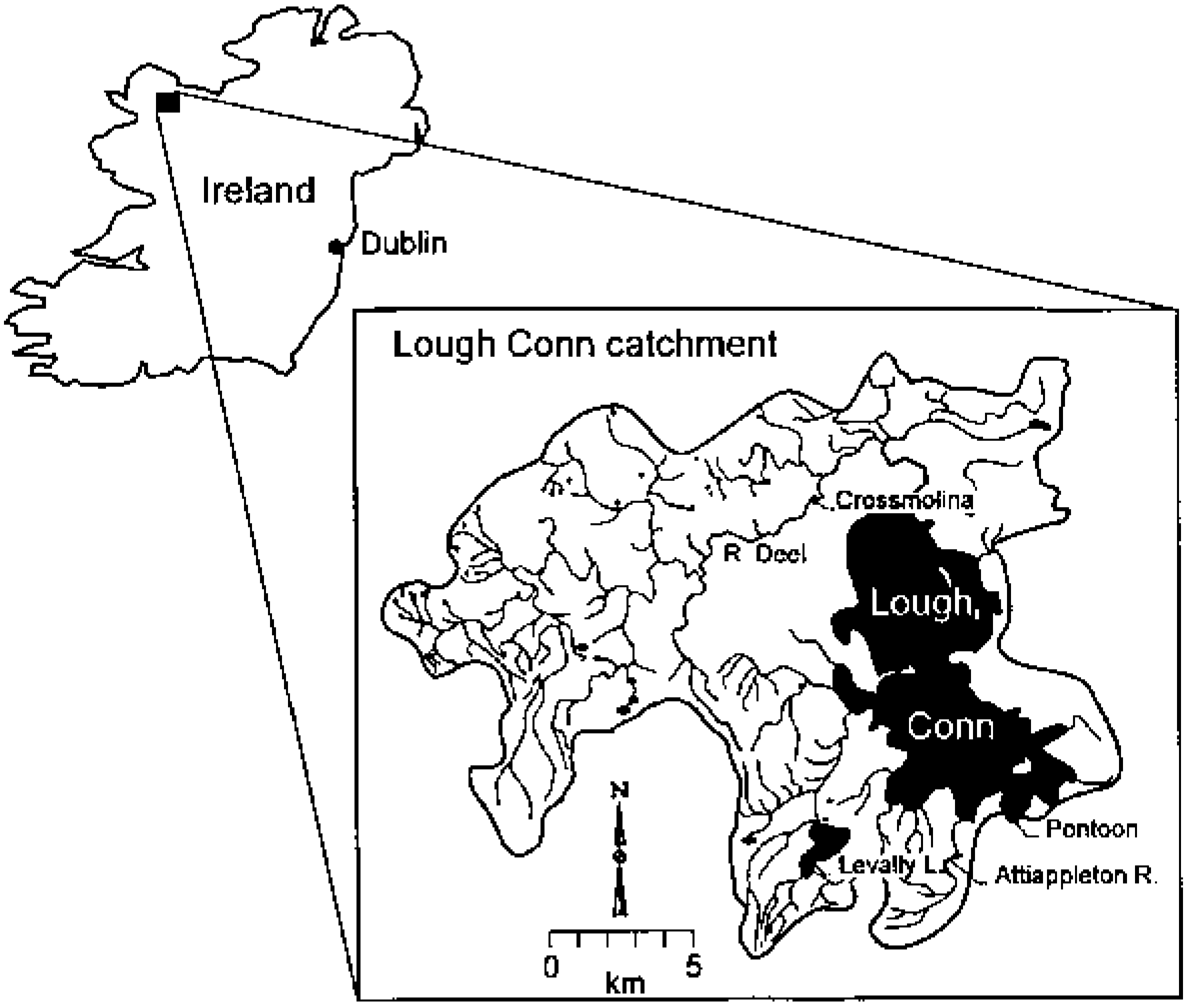
- *Analysis of results*: The statistical analysis of results should be described. The reliability of the statistics, and its implication, should also be given. Analytical procedures appropriate to the problem should be used and should focus on the goals of the study or the programme. Again, results should be presented in graphical, rather than in tabular, form.

- *Significance of results*: The report should include a section devoted to interpretation of the results in terms of the monitoring objectives. This will help ensure that the objectives are being met and that questions are being answered. This section should emphasise the need to provide information.

- *Recommendations*: Recommendations for proposed future activities should normally be listed in two categories: one concerned with scientific matters, the other with management issues. Ideally, these recommendations should be presented in order of priority.

- *Information sources*: All sources of information and all literature referred to should be correctly and fully cited so that a permanent record is available.

Figure 14.13 Example of a study area map (After McGarrigle *et al.* 1994. Reproduced with permission from Mayo County Council, Ireland)



14.7 Recommendations

It will be clear from the foregoing that the appropriate use and reporting of monitoring data is a vital part of the overall monitoring and assessment programme. Without clear reporting of understandable and relevant results to programme controllers (managers and administrators), little will have been achieved. That requirement can only be met if it is fully taken into account at the very earliest stages of the overall programme objective definition and subsequent design. It is also essential to understand that at every stage of the monitoring process, the data needs of the analysis and reporting stages are recognised. Without good data, no useful information may be reportable, no matter how good the underlying analysis may have been.

14.8 Source literature and further reading

Chapman, D. [Ed.] 1996 *Water Quality Assessments. A Guide to the Use of Biota, Sediments and Water in Environmental Monitoring*. 2nd edition, Chapman & Hall, London.

Cleveland, W.S. 1985 *The Elements of Graphing Data*. Wadsworth Advanced Books and Software, Monterey, CA.

Daniel, W.W. 1990 *Applied Non-Parametric Statistics*. 2nd edition. PWS-Kent Publishing Company, Boston.

Demayo, A. and Steel, A. 1996 Data handling and presentation. In: D. Chapman [Ed.] *Water Quality Assessments. A Guide to the Use of Biota, Sediments and Water in Environmental Monitoring*. 2nd edition, Chapman & Hall, London.

Elliott, J.M. 1977 *Some Methods for the Statistical Analysis of Samples of Benthic Invertebrates*. 2nd edition. Scientific Publication No. 25, Freshwater Biological Association, Ambleside.

Friedrich, G. and Viehweg, M. 1984 Recent developments of the phytoplankton and its activity in the Lower Rhine. *Verhandlungen des Internationale Vereinigen für theoretische und angewandre Limnologie,* **22**, 2029-2035.

Friedrich, G. and Viehweg, M. 1987 Measurement of chlorophyll-fluorescence within Rhine-monitoring - results and problems. *Archiv für Hydrobiologie Beihefte Ergebnisse der Limnologie,* **29**, 117-122.

Gilbert, R.O. 1987 *Statistical Methods for Environmental Pollution Monitoring*. Van Nostrand Reinhold Co., New York.

Humpesch, U.H. 1992 Ecosystem study Altenwörth: Impacts of a hydroelectric power-station on the River Danube in Austria. *Freshwater Forum*, **2**(1), 33-58.

McGarrigle, M.L., Champ, W.S.T., Norton, R., Larkin, P. and Moore, M. 1994 *The Trophic Status of Lough Conn. An Investigation into the Causes of Recent Accelerated Eutrophication*. Mayo County Council, Castlebar.

Pinder, L.C.V. 1989 Biological surveillance of chalk streams. *Freshwater Biological Association Fifty Seventh Annual Report*. Freshwater Biological Association, Ambleside, 81-92.

Snedecor, G.W. and Cochran, W.G. 1980 *Statistical Methods*. 7th edition. The Iowa State University Press, Iowa.

Sokal, R.R. and Rohlf, R.J. 1981 *Biometry*. 2nd edition. W.H. Freeman, San Francisco.

Spodniewska, I. 1974 The structure and production of phytoplankton in Mikolajskie Lake. *Ekologia Polska*, **22**(1), 65-106.

Williams, G.P. 1989 Sediment concentrations versus water discharge during single hydrological events in rivers. *Journal of Hydrology*, **III**, 89-106.