FRAMEWORK FOR DESIGNING SAMPLING PROGRAMS

W.A. MAHER, P.W. CULLEN and R.H. NORRIS
Water Research Centre, University of Canberra, P.O. Box 1, Belconnen ACT 2616, Australia

(Received: April 1993; revised: December 1993)

Abstract. A general framework for designing sampling programs is described. As part of the sampling program the problem of concern, or reason for sampling, needs to be clearly stated and objectives specified. The development of a conceptual model will assist the clarification of objectives and the choice of indicators to be sampled.

Objectives can then be stated as testable hypotheses and decisions made about the smallest difference/changes that are to be detected/observed by the sampling.

To allow the collection of representative samples, and the statistical analysis of data to be collected, the potential sources of variability in the data must be considered. Site, selection, frequency and replication must account for the expected variability.

Before field collection of samples occurs, the sample collection device needs to be tested as to its efficiency to collect a representative sample. It also will usually be necessary to consider how samples are to be preserved to inhibit biological and chemical change. All sample programs require a quality assurance program to identify, measure and control errors.

As well as the above the cost-effectiveness of the program should be evaluated in terms of maximizing the information obtained/cost.

1. Introduction

Considerable financial and human resources are used for sampling of the aquatic environment. In our view much of this is wasted through a poor standard of professional practice in the design and implementation of environmental surveys and monitoring programmes. It is hard to test this view with a formal evaluation since most such studies fail to specify any objectives in such a way that they can be measured whether the desired outcomes have been attained or not.

There are several Australian examples of massive collections of data which are of little use for the agencies that have invested public money in their collection and storage. They may be data rich, but they are often information poor. It is in this context of what we regard to be an inadequate standard of current professional practice that we offer this paper in an attempt to provide a better framework for the design of sampling programmes. We hope at least to stimulate a critical examination of current practice.

Not only is money being wasted but we believe that much of the data are unreliable for the uses to which they might be put. Managers and decision makers who take the trouble to extract the data and believe it, might well be seriously misled. As Horwitz and Howard (1979) suggest, the collection of samples that yield unreliable data are a potential cause of considerable social and political mischief. This situation is unsatisfactory since it means that water resources are being managed without the advantage of knowledge that could, and should, be available. Managers often pride themselves on the ability to make decisions with

Environmental Monitoring and Assessment **30:** 139–162, 1994.

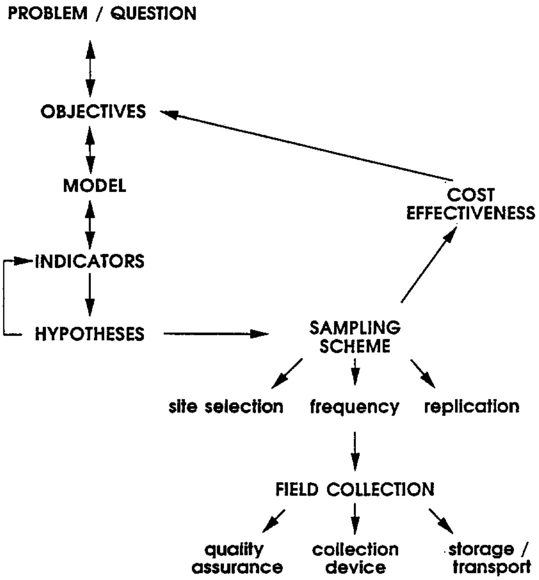

Fig. 1. Framework for designing sampling programme.

imperfect information. To make such decisions when they have paid for information that have not been supplied, or when the critical data could have been collected, but has not been, involves an unacceptable waste of public money, an unacceptable risk both to the water resource and to the users of the resource.

This paper proposes a framework for designing sampling programmes (Figure 1). The elements of the sampling design process are not sequential but interactive, so a holistic approach should be taken during the design phase and all of the elements considered.

2. The Concept of Sampling

Sampling is the collection from a defined population of a portion that represents the population as a whole concerning some variable. The problem of representation is obvious. Would we accept the height and weight of a single footballer selected randomly from the field as indicative of the average height and weight of the team? When we select non randomly, such as a footballer from the boundary, we may well exacerbate this problem.

Sampling presents an intellectual challenge to design a sampling approach that minimizes errors. Sampling can be thought of as an error-generating process (Gy, 1986). It must be assessed in terms of precision and accuracy. The aim of a sampling program is to collect useful information with the least cost. Data are not information, so if the samples cannot provide the information required they are not worth the time and expense of collection and analysis. Sampling refers both to the physical collection and removal of a subset of the system for later analysis; it also refers to taking an *in situ* measurement at a selected place and time.

Before a sampling program can be designed, the potential sources of variability in the data must be considered. Hypothesis testing relies on the falsification of a statement and statistical tests are designed to reject or accept the statement within certain predetermined levels of probability. We are only sampling part of the population of interest (the set of all possible measurements) and there will be an error associated with data collected about such a subset of the population. The amount of uncertainty that is acceptable depends on the intended use of the data. Variability in data may cause the rejection of a null hypothesis when it is actually true (Type I error) or acceptance of the null hypothesis when it is actually false (Type 2 error). Typical types of variation are:

- Spatial variability of indicators because of environmental heterogeneity,
- time dependency, temporal, seasonal effects,
- disruptive processes,
- dispersal of pollutants.

Environmental heterogeneity, both temporal and spatial, is probably the most significant aspect to be considered in the design of sampling programmes (Eberhardt, 1978; Morin *et al.*, 1987; Kerekes and Freedman, 1989). Variability will determine the number of sites, number of replicates and the frequency of collection. High environmental variability and logistical and financial constraints on sample collection and analysis often result in data too variable to detect an impact, disturbance or trend.

There are many questions in sampling that can be considered as ‘How’ questions: – how to collect, what to collect, when to collect, where to collect and how to store and analyse samples. These are unanswerable without a clear specification of the information required. Without knowing the answer to the ‘Why’ question it hardly matters how we answer the ‘How’ questions.

3. Clarifying the Problem and Establishing Objectives

3.1. THE PROBLEM TO BE ADDRESSED – THE WHY QUESTIONS

Before a sampling programme can be planned, the problem of concern or reason for sampling, needs to be clearly and unambiguously stated. There are many possibilities.

- Assuring that water quality is appropriate for particular uses.
- Assessing compliance with pollution control requirements.
- Identifying short term shock loads of pollutants and their impacts.

- Identifying long term trends in water quality.
- Reconnaissance survey of water quality for planning purposes.
- Testing hypotheses relating to structure and function of ecosystems.
- Investigating some undesirable conditions.

Our experience has been that managers are concerned with the observable symptoms in the water. Rather than treating the symptoms, scientists must identify the cause of the problem. Time and energy are required to formulate what the problem may be. It is not reasonable to assume that the manager has the scientific knowledge to do this. A professional approach will consist initially of identification and articulation of the problem, and this is an interactive process with the client.

How a problem is defined determines the approach taken to solve that problem. How a problem is defined will be a function of values, previous knowledge and experience. The initial statement of the problem may be the most crucial single factor in determining whether a solution can be found to the problem. Being able to redefine or reframe a problem and to explore the problem may broaden the range of alternatives and the solutions examined. Bardwell (1991) identified some pitfalls to be avoided in problem specification.

- Solving the wrong problem through not understanding the issues,
- stating the problem in a way that no solution will be possible,
- the premature acceptance of a possible solution before the problem is properly understood,
- use of information that is incorrect or irrelevant.

3.2. ARTICULATING OBJECTIVES

Clear objectives make it possible to design a sampling programme to obtain the information required. The problem is that many present practitioners seem to believe that some general sampling programme can answer any of their questions. This is not true. Many practitioners seem to prefer the activity of getting into the field and collecting some data, and hope they will be able to make sense of it later. It is common to find the wrong variables have been measured at the wrong place and time. Such sampling is very expensive, since there is no return on the investment.

In his guidelines for assessing scientific programmes, Smith (1985) states that objectives should:

- Be clearly and concisely defined,
- specify what is to be achieved,
- deal only with attainable results,
- not be expressed as idealistic aspirations,
- indicate when each stage will be completed.

The setting of objectives might involve only a scientific issue, but commonly the objectives will relate to management of a resource. This means the resource manager needs to be involved in the negotiation of objectives. The resource manager will often have only a limited range of options available, and will seek to specify objectives that improve the capacity to make an appropriate choice between them.

The resource manager needs to be clear about how the information to be collected will be used in the decision-making process. Managers and scientists need to interact to clarify the objectives for sampling programmes that are to support management. This will require agreement on the following:

- What are the important components of the system and what are the important linkages likely to be?
- What are the appropriate spatial and temporal boundaries for the study?
- What are valid indicators for the processes of concern?
- What precision is required in the data?
- What accuracy is needed in the data?
- Do criteria exist to help interpret the data, or is it necessary to answer questions such as the level of significance required at the design stage?
- Will the data to be collected be compatible with existing data, in terms of historical data collections, and in terms of related data such as hydrologic information?
- What sort of management options exist, and what data might support the analysis of various options in decision-making?

Interaction between the end users of information and the collectors of data must take place before sample collection starts. This design stage is fundamental to ensuring cost effective sampling programmes, and is often done poorly or not at all.

There are two critical decisions that require interaction between the collector and user of the information:

1. What is the smallest size of the change in each indicator that must be detected?
2. What certainty is necessary that changes of this magnitude are in fact detected by the sampling programme?

4. Sharing the Conceptual Model of the System

4.1. IMPORTANCE OF DIFFERING PERSPECTIVES

Our experience has been that professionals from different backgrounds have different conceptual models of a system. By conceptual models we are talking of simple box diagrams that show components and linkages. What are the driving

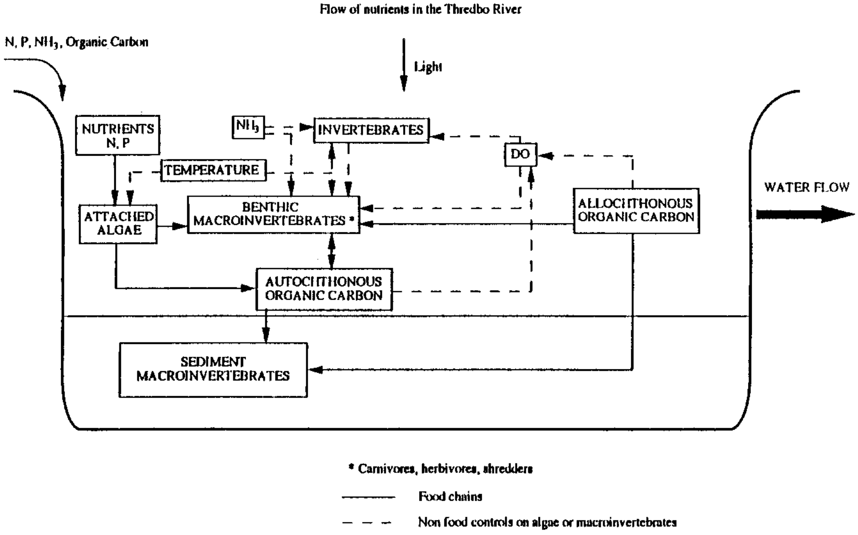

Fig. 2. Conceptual model of the Thredbo River.

factors, what are the consequential factors? We are not talking about numerical relationships or data-driven models at this stage.

An example of a conceptual model for the Thredbo River is shown in Figure 2. We were interested in assessing the impact of nutrients in sewage effluent discharged from the Thredbo Alpine Village on the river's water quality and ecosystem.

Our experiences on the Thredbo river and elsewhere show that people with backgrounds in primary production see problems differently from those concerned with secondary or higher level production. Take a fish biologist and an algal ecologist to a stream and listen to the problem analysis and decide if they are talking about the same stream. Add chemists and hydrologists to this team and you find an interesting diversity of models of how the system works.

These conceptual models are a powerful tool when we can argue about them and come to a shared model that satisfies us all. When this bargaining process is not undertaken, the different concepts of the problem can lead to disagreements about operational decisions and the importance of various types of data. Either people give in and don't care, which can lead to samples being collected which are not needed or used, or tedious arguments about how to do something drag on because the protagonists are working on different assumptions about how the system works.

It is essential for each team member to develop their concept of the system, and then to share and integrate these conceptual models. It should not be left to one team member, however experienced, since the differences in the models can be important in clarifying the real problem.

Limnology and oceanography are integrating sciences, and we need to exploit the problem-solving power of this integration. Once an appropriate model of the system has been made explicit and agreed, then many of the design questions become more obvious.

However, all models are a simplification of reality and involve personal judgment. The models do not need to be comprehensive, and embrace all components of the system; they only need to be adequate for the problem/question being investigated. The scientist will be aware of many other interesting questions, but will need to focus on answering the agreed question. The other questions will have to await other opportunities. The manager needs to accept a similar discipline. Broad scale, hypothesis-free data collection is rarely useful for anything, or is at least very inefficient.

At the same time it is essential to consider that the conceptual model being used might be wrong. Data that seem inconsistent can be important, leading to significant scientific breakthroughs when new and more powerful conceptual models can evolve.

4.2. SPATIAL BOUNDARIES AND SCALE

Once the conceptual model is agreed, the spatial boundaries of the system being investigated can be set and questions of scale considered. The setting of boundaries is important since inappropriate boundaries might focus the study away from important driving or consequential factors. The investigation of impacts in rivers for example will normally require the spatial boundaries to be those of the catchment.

What also needs to be addressed is scale. This refers to the units of size or time at which the system is observed. What is the appropriate level of resolution to answer the questions of concern? These decisions are often made on the basis of the skills of the investigator rather than an analysis of the problem. Different processes operate at various scales. Different scientific traditions focus on their appropriate scales. The effects of pollutants such as trace metals on organisms can be observed at scales ranging from the sub-cellular to community organization (Table I). The selection of the best scale to measure some phenomena should be driven by the problem to be addressed, not the resources or the intellectual traditions of a particular discipline or laboratory.

Decisions of scale should be made after considering the measurement opportunities at their various possibilities and the likelihood of collecting reliable and valid measurements. The costs of data collection at the various scales need also to be considered, as should uniformity over space of the indicators of interest. The larger the spatial extent of data collection the greater will be the heterogeneity or the patchiness of the measures, and the greater will be the number of replicate samples required to achieve the same confidence in the results. It is essential to choose an appropriate scale relative to the phenomena under consideration and then sample at that scale.

TABLE I

Effects of Scale on assessing the impact of trace metals on organisms.

Scale	Impact
Cellular	Induction of mixed function oxidases, metallothionens histopathological changes
Individual	Reproduction, survival, behaviour, growth rate
Population	Age structure, size classes, abundance, immigration, emigration, growth rate
Community	Species richness, biomass, structure, function

4.3. LENGTH OF STUDY

The appropriate length of the study is an important issue. Few hydrologists are expected to make definitive statements on the quantity of water resources with data sets as short as 2 or 3 years, yet in the water quality field such expectations are common. What is a reasonable period for the study in which a sufficient variety of rainfall events (from droughts to floods) can be experienced to allow the investigator the opportunity to study the system under stress? If sufficient time is not available, what tools exist to extrapolate from a limited range of conditions to the spectrum of hydrologic events that may be experienced?

It might be that quantitative models for the type of problem of concern are available. These provide mathematical representations of systems interactions based on mass flux or some other basis. They are often used to indicate the fate of contaminants in systems (e.g. MacKay, 1990). The advantage of quantitative models is that the important processes are defined and sampling programmes can be designed to ensure critical driving factors or rate variables are measured.

5. What Indicators are Appropriate?

There are two considerations in selecting what are appropriate indicators or variables to be measured in any particular study.

1. Relevance to the conceptual model of the system of interest.
2. Feasibility of measurement with acceptable precision, accuracy and cost.

Many studies include measurements of indicators that do not relate to the conceptual model of the system on which the study is based, and for which no predictive power is assumed. It is hard to understand why these are included.

This latter point is important since there are many indicators that might be very useful in managing particular problems but cannot be measured. Dissolved phosphorus might be useful in managing algal bloom problems, but samples need to be filtered within minutes of collection if reliable results are to be obtained (Lambert *et al.*, 1992). This is often not feasible in field situations, and so total phosphorus is used instead. It may be less useful as a diagnostic feature, but it is simpler to collect and is widely considered as being adequate given the difficulties of obtaining reliable data on dissolved phosphorus.

There are often decisions to be made as to whether driving or causal factors (such as phosphorus) should be measured, or whether consequential or resultant factors (such as algal biomass or chlorophyll) are more appropriate to answer the question of concern.

6. Study Design Considerations

6.1. ESTABLISHING HYPOTHESES TO TEST

Specific objectives need to be stated in terms of testable hypotheses, and statistical tools will often be used in testing these hypotheses. Hypotheses usually take one of two forms:

1. Variable A in a specified area or over a given time does not differ from that of a given base line by more than some predefined difference.
2. Variable A in a specified area not changed by more than some predefined difference per defined unit of time.

Hypotheses must be written such that two outcomes are possible – either rejection or acceptance. The null hypothesis (there is not a significant difference) can never be proved to be correct but can be rejected with known risks of doing so by using statistical power analysis (Fairweather, 1991). Any assumptions made in establishing hypotheses need to be stated as their validity must be examined as part of the sampling design.

7. The Pilot Study

Some idea of the variability in the system being sampled may be gained from examining published work although normally it is necessary to undertake a pilot study. This will give some idea of the underlying temporal and spatial processes and the sampling programme can be modified in light of the observed processes.

Tukey (1977) has proposed several graphical or numerical techniques to discover important patterns and statistical characteristics.

More recently, Monte Carlo simulations have been used in which the underlying temporal and spatial processes are simulated and statistical tests applied to calculate the uncertainty based on various sample sizes (Green, 1979). Many statistical techniques exist which deal with heterogeneity, e.g. ANOVAs with multiple comparison techniques, ordination analysis and multivariate techniques. The analysis used will depend on the information required. However, violations of assumptions, e.g. non-normally distributed data, non-independent data, dependence of variance on mean and the applicability of the statistical test must be established. Many statistical packages, such as Statistical Analysis System (SAS) can examine scatterplots and histograms of residual/errors for evidence of violations. It may not be important that statistical assumptions are met exactly but it is important to understand the importance consequences of any violations, and the effect these may have on the outcomes of the tests.

8. What Components of the Ecosystem should be Sampled?

The conceptual model of the system, and an understanding of pathways, is important in making decisions about where to sample the system. For example, some organic chemicals may be unmeasurable in the water itself, but might be detectable in sediment or in the biota.

The position of sampling within the site is also important; for example, it has been shown that trace metal concentrations of intertidal or sublittoral organisms can vary significantly (Nielsen, 1974; Phillips, 1976; Phillips, 1980).

Another interesting example of these locational problems comes with the measurement of chlorophyll A. Chlorophyll is rarely uniformly distributed with depth through the water column. It is commonly in bands, and often the water in the top metre is quite unrepresentative of chlorophyll concentrations. During blooms, most of the chlorophyll might be in a scum on the surface, which can move around depending on prevailing winds. We do not have good techniques for sampling this horizontal variability, and often just report the presence of a scum. In the vertical dimension it is common to take an integrated sample with a tube. This averages the chlorophyll over the depth and gives a mean concentration that might not be found anywhere in the water column. The depth to which the tube sample is taken is also an important and often unreported variable. Common depths are to 2 m, 5 m or to the depth of the euphotic zone.

9. The Spatial Selection of Sampling Sites

Site selection will depend on the study objectives and what is being studied. Sampling sites are usually selected by personal judgment using pragmatic considerations such as accessibility and safety. When impacts are being assessed sites will normally be located relative to the likely impact. Only rarely will sites be located randomly, but when this is done the number of sites and the extent of homogenous areas in which they may be located can be determined from a pilot study. Multivariate classification procedures can be used for grouping like sites to define homogenous areas (Norris and Georges, 1993).

Site selection in rivers downstream of a contaminant input needs to be validated to ensure adequate mixing takes place at a range of flows. Adjacent sites may yield samples in which the variables are spatially correlated. Semi variograms (Flatman *et al.*, 1988) can be used to quantify spatial correlation and to determine how far apart sampling sites must be to be independent. Autocorrelation analysis can also be used to test for independence between sites (Norris and Georges, 1993).

Sites need to be accurately located to allow repeat sampling, under a variety of likely future conditions.

Contaminants in sediment might not be uniformly distributed, but high near sources and low elsewhere. Is it appropriate to sample just one of these places? It is clearly inappropriate to sample both and take the mean. These decisions are easier if clear and specific objectives have been established, and appropriate methods selected (for example, spatial gradient analysis in this case). Systematic sampling, where samples are collected at regular intervals in space or time, is often used. Sampling sites are selected by personal judgment to best cover the area and may be biased. If this type of sampling is chosen, assumptions need to be stated and choices validated to prevent criticisms.

Random sampling is a requirement of many statistical tests. There are clear procedures for achieving this (Cochran, 1977), and they are not based on haphazard sampling, or whim on the day.

A substantial reduction of variability can often be achieved by replacing random sampling with stratified random sampling in which the system to be sampled is divided into parts (strata) each as uniform in the variable of interest as possible. Strata do not need to be of equal size and the number of samples is usually in proportion to the variance of the strata. For example, a lake can be divided into two strata (epilimnion, hypolimnion) for water sampling to obtain nutrient, chlorophyll and algal measurements. If we are collecting fish in a lake to look at the accumulation of pollutants, pollutant concentration in fish often increase with age. Fish may also be mobile. Fish age (size) becomes the sampling strata not geographical location.

Stratified sampling is judgmental in that prior information is used to choose strata but is probably the best compromise between random and systematic sampling as it is relatively free of personal judgment and reduces replication needs. Sampling precision is improved because uncertainty arises from variations within strata not differences between strata.

Green (1979) identifies three types of study design for detecting changes.

1. Optimal design for detection of change (before, after, control, impact or BACI).
   Replicated colletions made at more than one site in each of the control and impacted areas before the impact has occurred. The same sampling after the impact has occurred. A test of hypotheses based on this design will confirm whether a change has occurred and the magnitude of the change, within some level of predetermined confidence. It is rare to be able to implement such a design.
2. Inference from change over time.
   In this design there is no control site without any impact and change is established by comparison of data from one or more sites before and after the event being investigated. In this design there will always be uncertainty as to whether change might have occurred naturally over time independently of the impact.
3. Inference from change over space.
   In this design the control area with no impact is assumed to be upstream of some activity, and sites are located in the zones thought to have changed and those thought not to have been affected. Recovery zones and natural downstream changes may occur and these will add to the variability of the data. In such studies uncertainty exists as to whether the upstream and downstream sites may have differed before the impact.

Underwood (1991) has demonstrated that these approaches are often difficult to implement and suggests that changes in variability before and after impact may be of more use. Many impacts do no change the long run mean of a natural population but may change the temporal variance in the abundance of the population.

10. The Timing and Frequency of Sampling

Should sampling be done on a calendar basis such as daily, weekly, monthly, quarterly or some other period or basis? Are seasons important? What is the basis for making such decisions?

Some indicators give snapshots of immediate condition; some are integrating measures that reflect conditions over the past x months. These time scale decisions need to consider:

- the characteristics of the indicator being measured;
- the purpose of the data collection;
- the statistical or other tools that will be used to interpret the data. For instance, time series analysis may require a set sampling interval, and the critical decision is what the interval should be;
- the characteristics of the response of interest. For example, weekly measurements might be appropriate during the development of an algal bloom; they

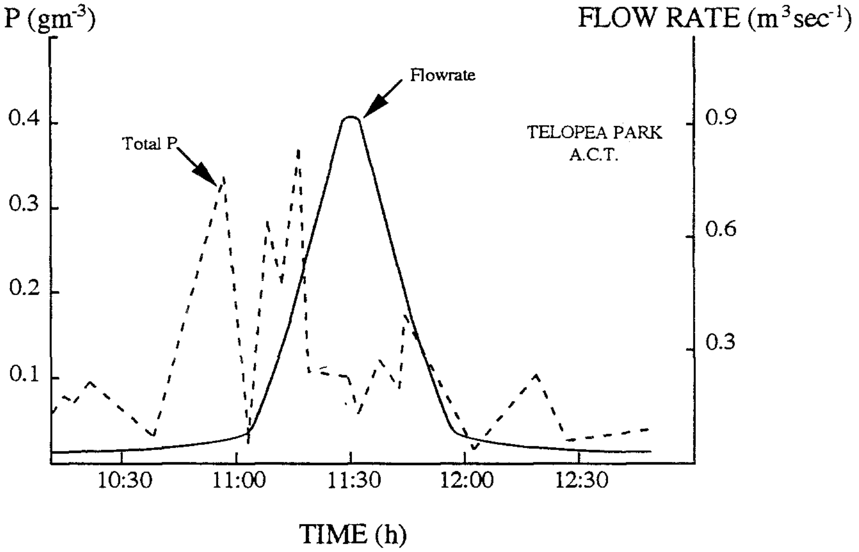

Fig. 3. Phosphorus concentration and discharge during a storm event.

would not be appropriate for investigating fish. The generation time of the organism might be the critical determinant of time scales;

- anything that takes longer to happen than the period over which measurements are made cannot be detected.

Some phenomena, such as the mass transport of substances, are best sampled on some hydrologic basis rather than on a calendar basis. Runoff events transport particulate matter and substances like nutrients and agricultural chemicals into streams. Higher flows resuspend material that had settled out. There may also be seasonal variations relating to grass cover and agricultural land management in the catchment that affect the quantity and quality of the runoff.

Some indicators will be highly related to flow. For example, non-point sources of phosphorus (Figure 3). Much of the total annual mass transport or load will be moved during a short period of high flows during flood events. For these measurements it is important to sample during high flow events and large numbers of measurements taken during low flow may be relatively unimportant. Event-based sampling is best undertaken using automatic sample collection equipment that is activated by changes in stream height. Decisions have to be made as to whether to use continuous data collection or sampling.

The frequency of sampling is especially important when monitoring to ensure a particular criterion or standard is not exceeded. Figure 4 shows some possible interpretations that might come from sampling at inadequate frequencies.

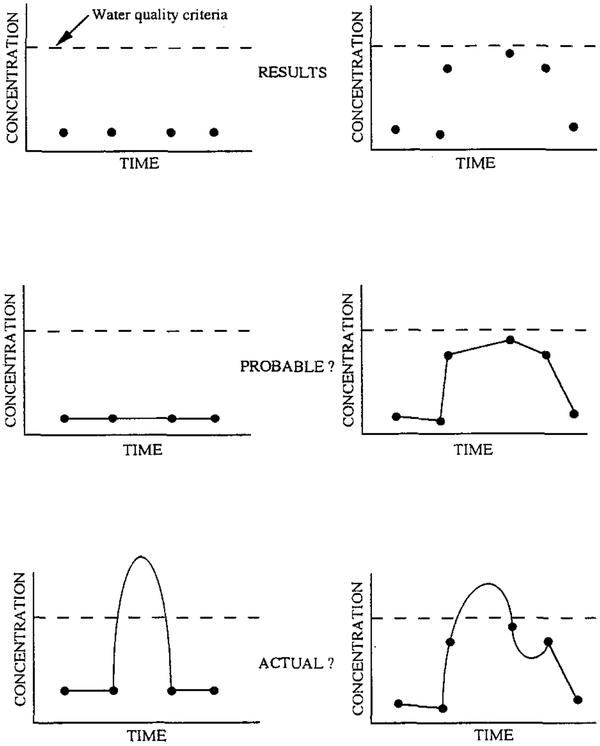

Fig. 4. Frequency of sampling: interpretation of sampling data.

It should be noted that samples taken close together in time (as well as space) are often correlated and not independent. Thus normal parametric statistical procedures to interpret data should not be used. As stated earlier, autocorrelation can be tested using a simple correlation of a variable with itself and preceding values. A time series with no serial dependance will produce correlation coefficients scattered around zero (range +1 to −1). Most statistical packages can perform such tests.

Mathematical formulae exist to calculate the sampling frequency required for a particular study (Montgomery and Hart, 1974) but are not in widespread use.

Biological sampling must also take into account the time dependency of organisms behaviour. Magman (1991) re-examined a published study on the Northern Red Belly *Phoxinus eos* and *Phoxinus neogaeus* in which the densities of both fish were reported as being highest at or near shore. The reported conclusion was that both species exploit the same microhabitat. Fish were sampled by trapping over a 16–18 h period beginning at 1600–1900 h. It was not recognized that *Phoxinus eos* have a diurnal pattern of inshore–offshore migratory behaviour. The fish swim in shoals in the inshore zone (< 0.5 m) depth during the day and migrate to the offshore zone (< 2 m depth) at sunset when shoals break up into single fish then go back to inshore zone at sunrise. A shorter interval of sampling (3–4 h) was required to observe this movement. The density of fish offshore seemed to be lower as the fish shoals had broken up. Subsequent diet studies revealed that the *Phoxinus eos* diet was zooplankton rather than green alge or diatoms, indicating the fishes' main food source was offshore not in-shore.

11. Selecting Appropriate Statistical Tools

11.1. SPECIFYING NEEDS

The statistical analysis to be used will depend on the information required. Different statistical procedures have different data requirements, so these decisions need to be made before data collection starts. Calling in a statistician and hoping that a flawed sampling programme can be fixed after the event is not acceptable, since frequently the appropriate analysis will be impossible.

Once the objectives are clear, issues like accuracy and precision can be addressed. There are tradeoffs with costs, but sampling in a way that does not enable the question to be addressed is a total waste of resources.

Decisions must be made about the smallest differences or changes that are to be detected, since these judgements determine the number of replicates needed (Norris and Georges, 1986; Norris *et al.*, 1992). If a phosphorus water quality standard of 50 mg m^{-3} exists, is it important to be able to identify 50.1, 51, 55 or whatever? This is the issue of precision. Decisions must be made on the precision needed for any estimate. Formulae are available for calculating the required replicates for each case (e.g. Norris *et al.*, 1992). Precision estimates can be calculated when it is desired to know the concentration of substances or the population number of organisms within desired limits. Frequently though, the levels of variables will not be as much interest as the differences or changes in them.

The establishment of the appropriate level of resolution is not a simple task (Segar *et al.*, 1987) since it must be:

- scientifically attainable;
- attainable through a sampling and analysis programme which can be accomplished in a cost-effective manner;
- environmentally significant in that change must represent something meaningful to the system of concern.

Once the difficult scientific questions of what level of precision is required, and the size of the differences that must be detected, are answered then the statistical question of how many replicates is required can be answered by performing the appropriate calculations (Norris *et al.*, 1992).

12. The Issue of Replication

Statistical techniques such as 'ANOVA' rely on the observed differences in measurements between sites or times being greater than the observed variation within sites or times (Sokal and Rohlf, 1981). The demonstration of a statistically significant difference will require replicate collection. Failure to include adequate replication in a sampling programme will lead to uninterpretable results. Alternatively, too much replication will be a costly waste of resources. Approaches to establishing the number of replicates required to yield the desired level of precision and resolution have been discussed in Norris *et al.* (1992).

Normally samples within a site are collected randomly such that the sample (variable) has an equal chance of representing the whole. An equal chance of being selected during sampling is a precondition for valid statistical conclusions to be drawn. There should be no conscious or unconscious selection of samples. Samples selected in a casual or haphazard way are not random. Random number tables or grids with random orientation of axes can be used. Because of the inherent variability in natural systems random sampling will require the greatest replication.

13. The Collection Device

If samples are to be collected for later analysis, the sampling device to be used will need to be tested as to its efficiency to collect a quantitative representative sample without disturbing the environment being sampled. Sample contamination may also occur by the device being in contact with media other than the sample of interest.

Green (1979) in his ten principles of sampling stated the need to 'verify that the sampling device is sampling the population you think it is sampling with equal or adequate efficiency over the entire range'. This requires specification of what population is to be sampled and what is the likely spatial and temporal variability. The ability of the collecting device to collect an undisturbed and representative sample might need to be tested. Device-related sampling errors cannot be accounted for by statistical methods or replication, and in many cases they will be undetectable unless specific tests have been undertaken. In an Australian river situation discharge can change by two orders of magnitude, and the effectiveness of sampling devices may change over this velocity range.

The sampling device should not significantly disturb the environment being sampled or alter the samples taken since such disturbance will mean samples do not reflect what 'was' or 'is'. The problems in sampling sediment illustrate these difficulties. Blomqvist (1991) has reviewed the problems of using several types of grab samplers and coring devices to obtain sediment samples. Grab samples often do not enter sediments perpendicularly and mixing of sediment layers occurs on closing. Most grabs have jaws which close semi-circularly and sediment layers below the initial penetration are only semi-quantitatively sampled. For quantitative sampling it is necessary to know the area and depth sampled. Coring devices require an unimpeded water flow or easily resuspended surficial material will be washed away. If rotation of cores occurs shear stress may mix the sediment and cause core shortening.

Quantitative biological sampling also presents a challenge. If trawling is used to catch fish, the question arises: are you capturing a representative sample of fish? Fish may be avoiding the nets and only particular species and sizes (ages?) of fish may be being caught by the trawl. Devries and Stein (1991) in their comparison of the efficiency of three devices (tube sampler, vertical tow net, Schindler-Patlalas trap) for collecting zooplankton found there was no best method. Zooplankton consists of a mixture of copepods, cladocerans and rotifiers. Generally copepods and cladocerans were best collected using the tube sampler while rotifers were best collected using the Schindler-Patalas trap. However, some species were best collected by the vertical tow net.

Some consideration must also be given to the environment traversed by the sampling device or sampling errors may occur by the device being in contact with media other than the sample of interest. For example, when collecting sub-surface water samples for hydrocarbon analysis the sample collection device must enter the water closed because as it passes through surface microlayer it will pick up hydrocarbons.

Sampling devices should be tested under controlled conditions to determine if the device quantitatively collects the sample of interest. In lieu of this, studies that have been reported in the literature that compare the efficiency of sampling devices and document the limitations of various alternatives (e.g. water samplers: Harris and Keffer, 1974; sediments: Blomqvist, 1991; Schneider and Wyllie, 1991) should be consulted. Using this information, a choice of sampling device can be made based on the matrix to be sampled and the unique conditions at the chosen sampling site.

14. Storage and Preservation

Once the samples have been collected it will usually be necessary to preserve them to retard biological, chemical and physical changes. Preservation choices will vary depending on the indicator to be measured. Some possible changes and preservation/storage procedures are given in Table II.

Considerations for preservation and storage include selection and decontamination of sample containers, selection of a preservation technique and the time lapse acceptable between sample collection and analysis. Choices will vary depending on the variable to be measured. Standards exist to provide guidance in this area (e.g. Australian Standard 2031.1 – 1986 ‘Selection of containers and preservation of water samples for chemical and microbiological analysis’).

TABLE II
Possible changes in samples and preservation techniques.

Change	Preservation techniques
Physical	
Adsorption/absorption	Inorganic: reduce pH on storage; organic: add solvent
Volatilization	No head space
Diffusion	Choose correct container type and cap liners
Chemical	
Photochemical	Use dark containers
Precipitation	Lower pH, avoid use of chemicals which cause precipitation (e.g. sulphates)
Other	Add fixing agent
Biological	
Microbiological	Reduce pH, add bactericide, freeze
Cell degradation	Fixing agent

15. Managing Errors

Horivtz (1978) has proposed a total survey design concept to evaluate the amount of data uncertainty. An attempt is made to minimize the total error of the estimate of interest by controlling the magnitude of the individual error components. Possible sources of error in the sampling and analysis process are shown in Figure 5. Often emphasis is placed on minimizing laboratory analytical errors with little consideration of the larger sampling errors. All errors contributing to the total error need to be assessed before decisions are made as to which errors are to be reduced. Youden (1967) proposed that once an error component has been reduced to less than one-third of the total error it is not cost effective to try to reduce it further.

15.1. QUALITY ASSURANCE

All sampling programmes require a quality assurance programme. The aim of this is to identify, measure and control errors. Major systematic errors to be avoided

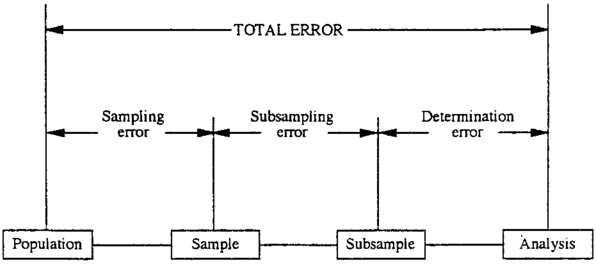

Fig. 5. Sources of error in sampling-analysis process.

are faulty sampling device operation, sample changes before measurement (e.g. contamination, chemical/biological changes) and incorrect sample labelling.

Where the possibility exists of the introduction of contamination into the sampling process a blank should be devised to detect and measure the contaminant. Field blanks in which a simulated sample is taken to the site and the container opened and closed and stored as for real samples during sample handling, transfer and storage are used to detect atmospheric fallout and other contamination. In freshwater work, sample bottles filled with distilled deionized water would be used as a field blank.

Equipment blanks in which the water/solvent used to rinse the sampling equipment between samples are used to determine contamination introduced through contact with sampling equipment. Trip blanks in which samples similar to sample collected but the analyte of interest is at background or low levels can be used to assess gross cross contamination of samples during transportation and storage. Control charts are used to detect changes in blanks (Lewis, 1988). Acceptable limits of change are set based on previous experience. Often we are not able to achieve no contamination but seek contamination levels to be stable. Contamination levels outside our acceptable limits indicate new sources of contamination.

Spiking of sub-samples in the field with a known amount of the analyte of interest and subsequent measurement will allow the detection of change. Quality assurance is a reactive process. If changes in samples are detected by spikes and blanks a specified procedure is required to determine and rectify the problem and resample if necessary.

15.2. DOCUMENTING FIELD METHODS

Sampling errors can be minimized by ensuring that correct procedures have been followed during field sampling, transport and storage. Detailed procedures to be followed in the collection, labelling, storage, transport and storage of samples and ancillary field data required need to be written and adhered to. Methods should be matrix and constituent specific and determine the sample collection device, type of storage container used and preservation procedures.

Training of the sampler to use sampling equipment is also specified within the methods. Attention should be given to anticipating problems in the field. Sample containers may be lost, sample volumes may be low. Do we include foreign matter? What criteria do we set for rejecting foreign matter? What do we do if sites cannot be sampled? Chain of custody procedures, whether or not required externally, are necessary if sample integrity is to be defensible.

The methods will also specify the types and numbers of quality assurance samples to be taken. This will require consideration of the nature of errors to be assessed both systematically and at random and the accuracy desired. Sources of error include reaction with sample/sample container, contamination (field, sampling device containers), chemical and physical instability, biological changes.

The entire sampling process should be scrutinized to minimize systematic sources of error that cannot be accounted for by the quality assurance program.

15.3. THE CONCEPT OF INTEGRITY

The quality assurance programme needs to ensure the integrity of the sample from collection to final analysis with respect to the variables of interest. If samples are to be the basis for legal proceedings at some time in the future, the following areas are likely to be under challenge.

- Exactly where was the sample taken from?
- Was the person taking the sample competent to do so?
- How was it labelled to ensure no possibility of mix up or substitution?
- Was there any possibility of contamination – of the container, of the sample during filling or later?
- Did the sample deteriorate after collection?

16. Cost Effectiveness

It is desirable that the cost of sampling programmes be minimized to meet the stated objectives. Cost-effectiveness considerations involve trade-offs between statistical power in the probability of discriminating between various hypotheses and the cost of data acquisition. It is necessary to determine all the resources and associated costs required to ensure the study can be carried out. Costs of data acquisition are determined by the:

- number of sampling stations;
- number of sampling occasions;
- replication;
- cost of collecting samples (staff, transport, consumables);
- cost of analysis;
- cost of data handling and interpretation.

TABLE III

Considerations when planning a sampling programme.

1. Has the problem/reason for sampling been clearly stated?

2. Are specific objectives:
   - (a) Clear and concisely defined?
   - (b) Sufficient to specify what is to be achieved?
   - (c) Specific enough to indicate when each stage is complete?
   - (d) Agreed between the users of data and the collectors?

3. Has a conceptual model of the system been made explicit and agreed?
   - (a) Have the study boundaries been agreed?
   - (b) Has the length of study been agreed?
   - (c) Has the scale of the study been agreed?

4. Have appropriate indicators been identified?

5. Have testable hypotheses been established?
   - (a) Will data from different sources be compatable?
   - (b) Will data collected yield information to test the hypotheses?
   - (c) Are statistical procedures clearly identified?
   - (d) Are the assumptions of the proposed statistical tests met?
   - (e) Has the smallest differences to be detected been specified?

6. Have the potential sources of variability been identified?
   - (a) Are there sufficient stations to accomodate variability?
   - (b) Is replication adequate to obtain the desired level of precision in data?
   - (c) On what basis is frequency of sampling proposed?

7. Will the sampling device collect a representative sample?
   - (a) Does disturbance of the environment being sampled occur?
   - (b) Does alteration of the sample occur by contact with the sampling device?
   - (c) What are the effects of the sampling device being in contact with media other than the sample of interest?

8. What programme is in place to identify, measure and control errors?
   - (a) How are samples to be preserved before analysis?
   - (b) Have sampling methods been written for samplers?
   - (c) Can the integrity of the sample be guaranteed?
   - (d) How are problems to be rectified?

Much information is available concerning the optimization of sampling programmes with regard to precision and cost (Eberhardt, 1976; Montgomery and Hart, 1974; Ellis and Lacy, 1980; Short, 1980; Bailey *et al.*, 1984; Hayes *et al.*, 1985; Lettenmaier *et al.*, 1984; Radford and West, 1986; Kratochvil, 1987).

Decisions will need to be made of what sampling effort is required to test critical hypotheses. If precision is reduced below that at which the critical hypotheses can be tested the proposed sampling design is a waste of time and money. If information is to be used to make decisions, priorities will often be based on the risks associated with making wrong decisions. Risk is often viewed not in environmental terms but political or social costs.

17. Conclusion

Management agencies have often underestimated the intellectual effort required to design and operate monitoring programmes, and have been unprofessional in their on-going scrutiny of the outputs of these programmes. Some scientists have also often not spent sufficient effort designing appropriate field sampling to enable hypotheses to be adequately tested.

We see sampling and monitoring issues to be an interesting science in their own right, and one where significant payoffs can be achieved. We have outlined a process which we believe leads to cost-effective sampling in that the data are useful for considering some specified question. The main points to be considered when designing a sampling programme are summarized in Table III.

If you cannot specify what is to be achieved, then sampling issues are hardly important. Measure what you like, when and where you like it. Don't expect these measurements to be interpretable and don't expect taxpayers to pay for them.

Professionalism involves helping the client understand the observed symptoms or phrase the critical question. It involves using state of the art physical and statistical tools to collect information that can be interpreted using a conceptual, deterministic or stochastic model of some sort. Professionalism also involves critical reflection on the whole sampling process to ensure cost effectiveness and to manage errors so they are kept within known and acceptable limits.

References

Bailey, D.A., Johnson, D., and Woolloff, D.A.: 1984, 'The Development of a Comprehensive Sampling Program for Yorkshire Water Authority', *J. Inst. Water Eng. Sci.* **38**, 435–456.

Bardwell, L.V.: 1991, 'Problem-framing: A Perspective on Environmental Problem Solving', *Environ. Management* **15**, 603–612.

Blomqvist, S.: 1991, 'Quantitative Sampling of Soft-bottom Sediments: Problems and Solutions', *Mar. Ecol. Prog. Ser.* **72**, 295–304.

Cochran, W.E.: 1977, *Sampling Techniques*, 3rd edn., John Wiley, New York, 428 pp.

Devries, D.R. and Stein, R.A.: 1991, 'Comparison of Three Zooplankton Samplers: A Taxon-specific Assessment', *J. Plankton Res.* **13**, 53–59.

Eberhardt, L.L.: 1976, 'Quantitative Ecology and Impact Assessment', *Journal Environ. Management* **4**, 213–217.

Eberhardt, L.L.: 1978, 'Appraising Variability in Population Studies', *Journal of Wildlife Management* **42**, 207–238.

Ellis, J.C. and Lacy, R.F.: 1980, 'Sampling: Defining the Task and Planning the Scheme', *Wat. Pollut. Control.* **79**, 452.

Fairweather, P.E.: 1991, 'Statistical Power and Design Requirements for Environmental Monitoring', *Aust. J. Marine Freshwat. Res.* **42**, 555–569.

Flatman, G.T., Englund, E.J., and Yfantis, A.A.: 1988, 'Geostastical Approaches to the Design of Sampling Regimes', in: *Principles of Environmental Sampling*, Keith, L.H. (ed.), ACS, Professional Reference book. American Chemical Society, pp. 73–84.

Green, R.H.: 1979, *Sampling Design and Statistical Methods for Environmental Biologists*, John Wiley and Sons, New York, 257 pp.

Gy: 1986, 'The Analytical and Economic Importance of Correctness in Sampling', *Anal. Chim. Acta* **190**, 13–23.

Harris, D.J. and Keffer, W.J.: 1974, *Wastewater Sampling Methodologies and Flow Measurement Techniques*, U.S. Environmental Protection Agency: Kansas City MO. EPA 907–974–005.

Hayes, C.R., Warn, A.E., and Green, L.A.: 1985, 'Development of Comprehensive Water Supply Quality Control in Anglian Water', *J. Inst. Wat. Eng. Sci.* **39**, 539–547.

Horivtz, D.G.: 1978, 'Some Design Issues in Sample Surveys', in: Namoodiri (ed.), *Survey Sampling and Measurement*, Academic Press, New York, pp. 3–11.

Horwitz, W. and Howard, J.W.: 1979, 'Sampling Methods for Trace Organic Analysis in Foods', NBS Special Publication 519. *Trace Organic Analysis: A new Frontier in Analytical Chemistry*, pp. 231–242.

Kerekes, J. and Freedman, B.: 1989, 'Seasonal Variation of Water Chemistry in Oligotrophic Streams and Rivers in Kejimkujik National Park, Novia Scotia', *Water Air Soil Pollut.* **46**, 131–144.

Kratochvil, B.: 1987, 'General Principles of Sampling', in: Taylor, J.K. (ed.), *Sampling and Calibration for Atmospheric Measurement ASTM STP 957*, American Society for Testing and Materials, Philadelphia, pp. 5–13.

Lambert, D., Maher, W., and Hogg, I.: 1992, 'Changes in Phosphorus Fractions During Storage of Lake Water', *Wat. Res.* **26**, 645–648.

Lettenmaier, D.P., Anderson, D.E., and Brenner, R.N.: 1984, 'Consolidation of a Stream Quality Monitoring Network', *Wat. Res. Bull.* **20**, 473–481.

Lewis, D.L.: 1988, 'Assessing and Controlling Sample Contamination', in: Keith, L.H. (ed.), *Principles of Environmental Sampling*, ACS Professional Reference Book American Chemical Society, pp. 130–144.

Mackay, D.: 1990, 'Environmental Data and Modelling', *Toxicol. and Environ. Chem.* **29**, 57–65.

Magmann, P.: 1991, 'Unrecognized Behaviour and Sampling Limitations Can Bias Field Data', *Environ. Biol. of Fishes* **31**, 403–406.

Montgomery, H.A.C. and Hart, I.C.: 1974, 'The Design of Sampling Programmes for Rivers and Effluents', *Wat. Pollut. Control* **73**, 77–101.

Morin, A., Mousseau, T.A., and Roff, D.A.: 1987, 'Accuracy and Precision of Secondary Production Estimates', *Limnol. and Oceanog.* **32**, 1342–1352.

Neilsen, S.A.: 1974, 'Vertical Concentration Gradient of Heavy Metals in Cultured Mussels', *NZ J. Marine Freshwat. Res.* **8**, 631–636.

Norris, R.H. and Georges, A.: 1986, 'Design and Analysis for Assessment of Water Quality, in: DeDeckker, P. and Williams, W.D. (eds.), *Limnology in Australia*, CSIRO Dr W. Junk, pp. 555–572.

Norris, R.H., McElravy, E.P., and Resh, V.H.: 1992, 'The Sampling Problem', in: Calow, P. and Petts, G.E. (eds.), *Rivers Handbook*, Blackwell Scientific Publishers, Oxford, pp. 1–15.

Norris, R.H. and Georges, A.: 1993, 'Analysis and Interpretation of Benthic Surveys', in: Rosenberg, D.M. and Resh, V.H. (eds.), *Biomonitoring and Freshwater Macroinvertebrates*, Chapman and Hall, pp. 234–286.

Phillips, D.J.H.: 1976, 'The Common Mussel *Myhlus edulis* as an Indication of Pollution by Zinc, Cadmium, Lead and Copper, Effects of Environmental Variables on Uptake of Metals', *Mar. Biol.* **38**, 59–69.

Phillips, D.J.H.: 1980, *Quantitative Aquatic Biological Indicators. Their use to Monitor Trace Metal and Organchlorine Pollution*, Applied Science, London, pp. 488.

Radford, P.J. and West, J.: 1986, 'Models to Minimize Monitoring', *Wat. Res.* **20**, 1059–1066.

Schneider, P. and Wyllie, S.J.: 1991, 'An Efficient Vibrocoring System for Collecting Coastal Sediments. A Comparison With Other Techniques', in: *Proceedings 'Coastal Engineering-Climate for Change'*, 10th Australasian Conference on Coastal and Ocean Engineering Auckland, 2–6 December 1991, pp. 357–362.

Segar, D.A., Phillips, D.J.H., and Stamnan, E.: 1987, 'Strategies for Long-term Pollution Monitoring of the Coastal Oceans', in: Boyle, T.P. (ed.), *New Approaches to Monitoring Aquatic Ecosystems ASTM STP 940*, American Society for Testing and Materials, Philadelphia, pp. 12–27.

Short, C.S.: 1980, 'Sampling Programme Design for Water Quality in Distribution', *Wat. Serv.* **84**, 529.

Smith, A.M.: 1985, 'Management Guidelines for Evaluating Scientific Activities in Departments and Declared Authorities in New South Wales', *Program Evaluation Bulletin*, Public Service Board of NSW Program Evaluation Unit.

Sokal, R.R. and Rohlf, F.J.: 1981, *Biometry*, 2nd edn., W.S.H. Freeman, London, 859 pp.

Tukey, J.W.: 1977, *Exploratory Data Analysis*, Addison-Wesley, Reading, Mas.

Underwood, A.J.: 1991, 'Beyond BACI: Experimental Designs for Detecting Human Environmental Impacts on Temporal Variations in Natural Populations', *Aust. J. Mar. Freshwat. Res.* **42**, 569–589.

Youden, W.J.: 1967, 'The Role of Statistics in Regulatory Work', *J. Assoc. Off. Anal. Chem.* **50**, 1007–1013.