GUIDELINES FOR WATER QUALITY MONITORING AND REPORTING

Barmuta LA[1]*, Maher WA[2] and Batley GE[3]
[1] School of Zoology and Tasmanian Institute of Aquaculture and Fisheries, University of Tasmania, GPO Box 252-05, Hobart, Tasmania 7001, Australia.
[2] CRC for Freshwater Ecology, University of Canberra, Canberra, ACT, 2601, Australia.
[3] Centre for Advanced Analytical Chemistry, CSIRO Energy Technology, Lucas Heights, NSW 2234, Australia.

ABSTRACT

The *Australian Guidelines for Water Quality Monitoring and Reporting* (ANZECC & ARMCANZ 2001b) is one of the series of documents that form part of the National Water Quality Management Strategy. It is complementary to the *Australian and New Zealand Guidelines for Fresh and Marine Water Quality* (ANZECC & ARMCANZ 2001a) and provides much of the generic, background advice necessary to plan and carry out a program of monitoring water quality. Emphasis is placed on clearly articulating the objectives of a program after a conceptual model of the key processes has been articulated by the monitoring team. This process will identify a number of candidate variables for potential monitoring which will include a range of physical, chemical, ecotoxicological and biological variables. The use of professional statistical expertise is then strongly encouraged when planning the study design of the program, which should clearly articulate the spatial and temporal features than need to be captured in sampling. Generic considerations of field sampling and laboratory analysis are covered, with extensive cross-references to protocols in the *Australian and New Zealand Guidelines for Fresh and Marine Water Quality* and elsewhere. A variety of descriptive and inferential statistical procedures can be used to explore and analyse the data, and, again, professional statistical advice is essential to ensure that methods are being chosen and used correctly. The method for using the trigger value procedure recommended by the *Australian and New Zealand Guidelines for Fresh and Marine Water Quality* for physical and chemical values is described in some detail. Finally, advice is provided about the reporting and dissemination of results to a variety of likely stakeholders in the outcomes of a water quality monitoring program.

Key words: water quality, monitoring, impact assessment, trigger value.

INTRODUCTION

Designing, carrying out and reporting a program of water quality monitoring or assessment requires a number of generic stages, irrespective of the nature of the indicators being used. The *Australian Guidelines for Water Quality Monitoring and Reporting* (ANZECC & ARMCANZ 2001b) (hereafter called *Monitoring and Reporting Guidelines*) sets out these stages of planning and execution of water quality monitoring programs, and provides background and support for the *Australian and New Zealand Guidelines for Fresh and Marine Water Quality* (ANZECC & ARMCANZ 2001a) (hereafter called the *Water Quality Guidelines*).

This paper outlines these generic stages, and describes in some detail the use of trigger values for physicochemical and toxicant indicators. These trigger values are a key part of the risk-based strategy of the new water quality guidelines, since exceeding the trigger value may initiate more comprehensive monitoring and/or management responses depending on the context of the program and the problem being addressed.

The standard structure for carrying out a water quality monitoring program is summarised in Figure 1. The steps involve: setting the objectives of the program so that the data requirements are clearly identified, designing the study including consultation with statistical professionals to ensure that the study will meet the objectives, implementation of appropriate field sampling programs and laboratory analyses with appropriate QA/QC procedures, analysis and

*Author for correspondence; email,

interpretation of the data, and reporting and disseminating the information gathered. As data are analysed, improvements to the program may be identified which should be fed back to the study design phase where appropriate. Each step of Figure 1 is accompanied by detailed flow charts and checklists of tasks in the *Monitoring and Reporting Guidelines*.

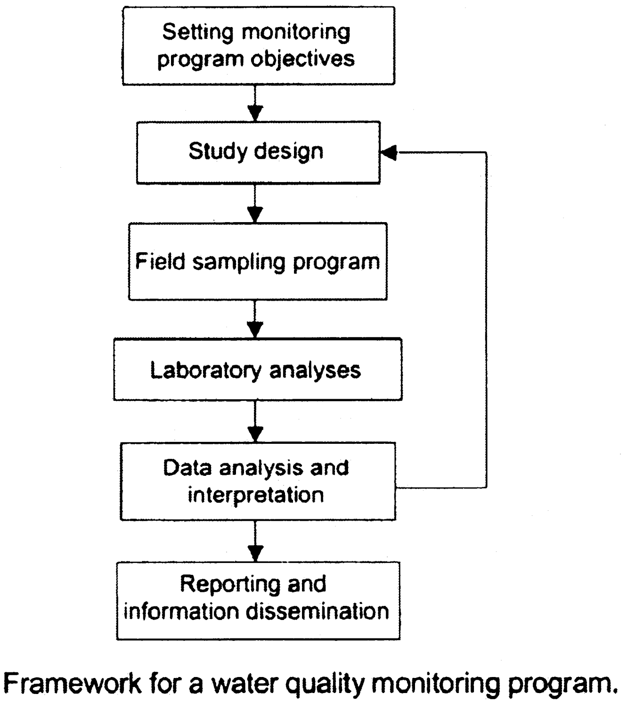

Figure 1. *Framework for a water quality monitoring program (after Figure 1.1 in ANZECC & ARMCANZ 2001b).*

SETTING OBJECTIVES

Monitoring of waters is typically undertaken to meet one or more of the following general objectives:

- measuring the ambient quality of water;
- assurance that the water meets guidelines for its designated use;
- investigation of why a water may not be meeting guidelines;
- assessment of loads of materials for mass balance studies;
- characterisation of the biota within a water body;
- assessment of biological productivity;
- assessment of the status of the water resource (eg. State of the Environment or National Audit reporting);
- assessment of the efficacy of management interventions; and/or
- identification of trends in the condition of the water body.

These general objectives need to be combined with a conceptual model of how the water body being monitored works. These conceptual models can be very simple diagrammatic representations (eg. Figure 2), but should result from collaboration between all members in a team conducting a monitoring or assessment program so that all of the features of the ecosystem and water quality issue under study are captured and worked through. As part of this process a suite of biological, chemical and physical variables will be identified as potentially useful in the monitoring program, including any pre-existing data about the water body. This information is then combined in an iterative fashion with the general objectives of the program to refine the specific objectives which need to be monitored or assessed. This refinement is crucial, and many water quality monitoring programs have failed in the past because the initial objectives of the program were not explicit enough or even achievable (Bardwell 1991).

The specific objectives that are identified at the end of this process will generally fall into one of two broad categories: objectives that can be framed as testable hypotheses, and objectives that are couched in terms of measuring the magnitude or trend of the chosen indicators of water quality. Which of these categories is applicable needs to be borne in mind when deciding how the data are to be collected (ie. the study design) and analysed and are issues that need to be canvassed with statistical professionals in the next phase, that of study design.

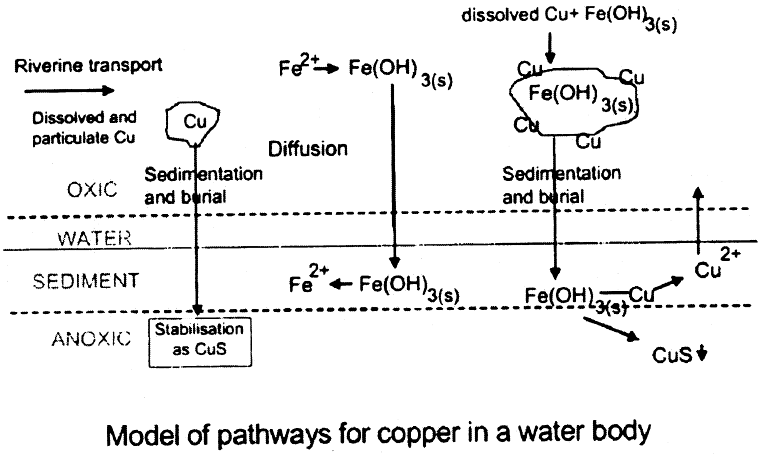

Figure 2. *An example of a conceptual model. This describes the pathways for copper in a water body (after Figure 2.5 in ANZECC & ARMCANZ 2001b).*

STUDY DESIGN

Types of study design

A program of monitoring water quality can fall into three broad categories of study design, and are introduced in more detail in Chapter 3 of the *Monitoring and Reporting Guidelines*:

• Descriptive studies, where the goal is to document the state of the system. Once the data have been collected it is not usually possible to analyse these data to demonstrate causality. Examples of descriptive studies include reconnaissance surveys, State of the Environment or Audit reporting and baseline studies (Green 1979;ANZECC & ARMCANZ 2001a).

• Studies that measure change. These studies usually involve sampling at more than one time and/or at more than one location. There are three main subcategories that can be identified: before-after, control-impact (BACI) designs and their derivatives; designs where inferences of change are based on changes over time; and, designs where inferences of change are based on spatial changes. These subcategories are outlined in Humphrey *et al.* (in press), and reviewed by Stewart-Oaten and Bence (2001).

• Studies that improve system understanding, usually by demonstrating cause and effect. If the objective of the program is to establish causality, the sampling program must be designed for this purpose from the start. This will often require one or more manipulative experiments, and in large investigations may involve a series of laboratory and field-based experiments. For complex phenomena, no program will be able to completely defend against all unidentified confounding influences, so a monitoring team must be able to assemble independent lines of evidence in much the same way that epidemiologists do. Beyers (1998) has adapted epidemiological criteria for such purposes, and these are summarised in Table 1.

Sampling in space and time

Once the objectives of the program and broad category of study design have been decided, the spatial boundaries of the study area need to be identified along with the duration of the study. As spatial and temporal scales increase, the phenomena of interest are likely to become more heterogenous. In addition, repeated measurements over time will need to accommodate time dependence, seasonal and interannual effects as the duration increases. The pattern of sampling, in both space and time, will therefore be crucial in capturing the features of the system that have been targeted by the monitoring program. Although most of us are familiar with simple random, stratified random and systematic sampling patterns, there are other variations on these themes which could be useful for certain applications (Thompson 1992). As ever, professional statistical advice at the planning stage of the study will be invaluable in deciding the most cost effective sampling regime.

Often the sampling program will involve several study sites or locations. In some programs some of these sites will represent 'control' or 'reference' conditions whereas one or more sites, termed 'test sites' in this paper and the *Monitoring and Reporting Guidelines* are those locations that where some management intervention or putative impact has taken place or may take place in the future. Where such arrangements are included in the study, care needs to be taken to ensure that the sites are closely matched, and that heterogeneity within the sites is properly accounted for in the sampling design. Sometimes covariates can be collected at all the sites to adjust the values of measured variables for inherent differences between the sites (Stewart-Oaten and Bence 2001).

A simple way of ensuring similarity between sites is to select those which are spatially close to each other. However, if sites are too close to each other, serial correlation between the sites can invalidate the assumptions of independence made in some classical statistical designs (Cressie 1993). What constitutes 'too close' depends both on the nature of the variable and its dispersion in the environment. The monitoring team should consider whether to select alternative sites, or if sufficient data can be collected to implement designs that can model these spatial patterns properly.

The issue of sampling frequency will, again, depend on the precise objectives of the monitoring program, nature of the variable being sampled, and its natural variation through time. If, for example, a variable has a predictable temporal pattern (eg. recruitment of a fish species with the onset of the wet season, or deoxygenation during thermal stratification), the sampling program must be frequent enough to suit this periodicity. If a disturbance is only likely to take place at a certain time of the year (eg. discharge of mine waste only during the wet season Humphrey *et al.* 1995) then sampling can be targeted to such predictable 'pulse' disturbances. At the other extreme, to measure the effects of highly variable and unpredictable disturbances (eg. stormwater discharges), the monitoring program must sample at several time scales. Some variables will give snapshots of immediate conditions, while others may integrate conditions over some extended previous time period. Thus these decisions about time scales need to be based on:

Table 1. *Criteria to formalise the use of independent lines of evidence in inferring causation in impact studies (from Beyers 1998). The example is an hypothetical one of the response of a biological variable (eg. population density) to a toxicant.*

Name of criterion	Description of criterion	Example
Strength of association	Size of the correlation between the intensity of the disturbance and the response of the measurement parameter.	Sites with high concentrations of the toxicant have lower population densities of an organism than sites with low concentrations of the toxicant.
Consistency of association	The association between the disturbance and the variable has been repeatedly observed in different places, circumstances, and times.	The negative correlation between concentrations of the toxicant and the densities of the organism has been demonstrated in several other studies by other investigators elsewhere.
Specificity of association	The observed effect is diagnostic of exposure to the disturbance.	In this case, decrease in the density of the organism is not diagnostic of the disturbance because the population density may be reduced by other, natural processes.
Presence of stressor in tissues	Measurement of variables of exposure (eg. residues, breakdown products) must be present in tissues of affected organisms.	Breakdown products of the toxicant are found in the tissues of organisms in sites with high exposure, but are below detection limits in sites where the toxicant is absent.
Timing	Exposure to the disturbance must precede the effect in time.	Accidental spillage of the toxicant are usually followed by sharp declines in the density of the organism.
Biological gradient	A dose-response relationship exists (ie. response of variable is a function of increases in magnitude of the disturbance).	Ecotoxicology tests have established a dose-response relationship.
Biological plausibility	There is a biologically plausible explanation for causality, even if the precise mechanism is unknown.	The toxicant comes from a group of chemical know to interfere with respiration in this organism.
Coherence	The causal interpretation should not seriously conflict with existing knowledge about the natural history of the organism and the behaviour of any substances associated with the disturbance.	The organism is usually common in sites within the study region and is present year-round; the toxicant is readily soluble and does not breakdown readily while in solution.
Experimental evidence	A valid experiment provides strong evidence of causation.	A field experiment demonstrated rapid mortality in response to the addition of known concentrations of the toxicant.
Analogy	Similar disturbances cause similar effects.	Other chemicals related to this toxicant have shown similar dose-response curves and responses in field experiment with different but related species.

- the characteristic of the variable being measured;
- the specific objective of the monitoring program;
- the statistical or other tools that will be used to interpret the data;
- the characteristics of the response of interest; and
- recognition that a process cannot be measured if it takes longer to happen than the period over which measurements are made.

A case study involving the Great Barrier Reef is described in an appendix of the *Monitoring and Reporting Guidelines* to illustrate these issues in more detail.

Each of the broad classes of indicators identified in the *Water Quality Guidelines* have some specific issues in temporal sampling that need emphasis. For biological indicators, periodic movement patterns (eg. diel variations in activity, or seasonal patterns in recruitment) and the longevity of the organism need to be accounted for in the sampling design.

Similar issues hold for chemical and physical variables (eg. diel patterns in dissolved oxygen in still waters, persistence of some contaminant over long time periods). Variables measured in the water column have additional issues relating to flow, especially if concentration measurements are being used to calculate loads. When flow is an important issue in the program, the following need to be considered:

- the importance of flow-based monitoring and of capturing first flush and peak events;
- the need to measure and record flow data in conjunction with analyte concentration data obtained at the same time; and
- the need to sample and obtain information at all flow regimes, including low flow, so that water quality can be described for all conditions of the water body.

Automatic sampling devices can be invaluable in collecting flow-related data.

Selection of variables

Through the process of developing a conceptual model and defining the objectives, the monitoring team will have a range of potential range of physical, chemical, ecotoxicological and biological variables from which to choose. No single or simple variable suffices to characterise water quality. The *Water Quality Guidelines* promotes the idea of integrated assessment where the biological consequences and physical and chemical causes of a problem are combined in the monitoring program. Thus a three-pronged or 'triad' approach, using chemistry, ecotoxicology and ecology has been advocated for comprehensive monitoring and assessment programs (Chapman 1990).

Choosing among the range of candidate variables is not a trivial process. Maher and Cullen (1997) provided six criteria which are outlined in Table 2. Trade-offs between different variables will be inevitable, and need to be made with these criteria in mind, together with information about the costs and difficulties involved in measuring the variables concerned. For example, the monitoring team's conceptual model may have identified the concentrations of bioavailable contaminants as key variables, but may settle for total concentrations because they are easier to measure and more reliable. Much more information about the different variables that can be used is provided in Chapter 3 the *Monitoring and Reporting Guidelines*, and the *Water Quality Guidelines* describes which variables are best suited to a wide variety of water quality monitoring issues, including details on applicable protocols (ANZECC & ARMCANZ 2001a).

Table 2. *Maher and Cullen's (1997) criteria for selecting variables.*

Relevance	Does the variable reflect the issue of concern?
Validity	Does the variable parameter respond to changes in the environment and have some explanatory power?
Diagnostic value	The variable must be able to detect changes and trends in conditions for the specified period. Can the amount of change be assessed quantitatively or qualitatively?
Responsiveness	Does the measurement parameter detect changes early enough to permit a management response, and will it reflect changes due to the manipulation by management?
Reliability	The measurement parameter should be measurable in a reliable, reproducible and cost-effective way.
Appropriateness	Is the measurement parameter appropriate for the time and spatial scales of the study?

FIELD AND LABORATORY PROCEDURES

The exact field and laboratory procedures that need to be followed depend on the objectives of the program and the study design that has been agreed on by the monitoring team. Specific protocols for various indicators are provided by the *Water Quality Guidelines* (ANZECC & ARMCANZ 2001a), and a number of other sources of specific information are referred to in Chapters 4 and 5 of the *Monitoring and Reporting Guidelines*. These sources need to be consulted to ensure that the acquisition, preservation and storage of samples is conducted in line with best current practice. Two additional generic aspects of field and laboratory procedures will be emphasised here: quality assurance and quality control (QA/QC) procedures and occupational health and safety issues.

A QA/QC program is intended to control sampling and analytical errors to levels acceptable to the user. This will include procedures designed to prevent, detect and correct problems in sampling and analytical processes, and will usually include procedures to characterise errors statistically through the use of quality control samples (eg. for chemical variables: the use of blanks, spiked samples, duplicated samples, etc.; for ecotoxicology: the use of negative controls, reference toxicants, etc.; for biological materials: independent re-checking of identifications, etc.). Field and laboratory staff should be competent in the use, calibration and maintenance of equipment and the deployment of sampling devices; where protocols are prescribed, they should be adhered to.

The tracking of samples via a documented 'chain of custody' is vital to ensure all activities relating to a sample are traceable. This chain of custody extends from proper and adequate labelling of material in the field through logging laboratory storage and analytical procedures and culminating in the appropriate storage of information in databases so that the information is readily retrieved. This information combined with the more familiar QA/QC procedures is vital if the data are used in legal proceedings.

Occupational Health and Safety issues are also important, with most jurisdictions in Australia having specific requirements for employers and employees to fulfil. For field work, procedures should be in place to ensure that hazards are identified, staff are sufficiently educated and informed about the hazards, and that options for minimising risks are taken. In some cases this may mean that alternative sampling sites need to be found, and this will need full consultation with the monitoring team. Furthermore, some programs will necessitate sampling water that is hazardous to human health (eg. water contains high concentrations of contaminants or pathogens), and appropriate protective clothing and sampling gear will be required. In the laboratory, similar procedures apply to the identification of and education about hazards and the development of risk minimisation plans to deal with them. The *Monitoring and Reporting Guidelines* cites several sources of information about health and safety procedures for chemical and biological laboratory facilities. Staff should not only be qualified in the handling of materials and equipment, but familiar with safety and first aid procedures.

DATA ANALYSIS AND THE USE OF TRIGGER VALUES

It is worth reiterating that the type of study design adopted critically affects the opportunities for statistical analysis and scope of any inferences or hypothesis tests that are made. Delaying involvement of professional statistical expertise until after data has been collected is foolish. That said, the scope of this section is to, first, highlight the variety of statistical procedures that may be relevant to water quality programs and second, to outline the method recommended for comparing physical and chemical indicators with trigger values as stipulated by the *Water Quality Guidelines*.

The variety of statistical tools

Chapter 6 of the *Monitoring and Reporting Guidelines* provides a more comprehensive introduction to this topic, with more technical details and worked examples provided in Appendix 5 of that document.

There is a variety of statistical schools of thought. The classical 'frequentist' methods that most us are familiar with have served us well, but there is increasing disquiet over their applicability to environmental situations where the data frequently violate key assumptions (McBride *et al.* 1993; Johnson 1999) and where 'reference' and 'control' conditions are not allocated at random (Stewart-Oaten and Bence 2001). Alternative statistical methods have been proposed, although some of them also have difficulties (Fox 1999). Again, the involvement of a professional statistician can guide users to the choice of appropriate statistical tools provided the objectives of the study are clearly stated and the design of it has provided data amenable to addressing those objectives.

As an initial step in data analysis, the *Monitoring and Reporting Guidelines* strongly advocate that users explore their data properly rather than blindly applying formulaic tests. These exploratory techniques include numerical summaries, data visualisation, transformations, detection of outliers, checking for censored data, trend detection and smoothing. Selection of which techniques to use depends on the objectives of the analysis and the nature of the data which have been collected.

Of these exploratory issues, censoring deserves further exploration because it is often a feature of chemical and some physical data sets. Some values of the chosen variable are below the level of detection ('below detection limit' or BDL) of the technique or instrument used to measure it. Such data are said to be censored, and censored data are especially problematic when a large proportion of the data are BDL. The common methods of treating BDL observations (eg. substituting zero or the detection limit, or coding them as 'missing') result in biased estimates when used in classical statistical techniques. A number of sophisticated techniques have been proposed for such censored data, but require professional expertise to apply properly. In the absence of this input and where only a small proportion of the data set is BDL, the analyses can be run twice: once with BLD values replaced with zero and once with the BDL values replaced with either the detection limit or half the detection limit. If the results of the two analyses differ markedly then more sophisticated methods of dealing with the censored observations need to be sought as described in Chapter 6 of the *Monitoring and Reporting Guidelines*.

For many biological indicators and for some programs involving physical and chemical indicators the *Water Quality Guidelines* (Section 3.2.4) recommends analysing the data within a hypothesis testing framework where data from the test site are compared with data from one or more reference sites. Further background information on these procedures is provided in Humphrey *et al.* (in press) and Chapter 7 of the *Water Quality Guidelines*. While much of the published literature in this area has focussed on the application of classical methods such as *t*-tests and ANOVA, the *Monitoring and Reporting Guidelines* draws attention to two more classes of statistical models: generalised linear models (McCullagh and Nelder 1983; Dobson 1990) and generalised additive models (GAMS) (Hastie and Tibshirani 1990). These have considerable advantages over classical methods in that they are more flexible than classical methods. For generalised linear models, a variety of error distributions can be chosen from (ie. the data do not have to conform to the normal probability model), and non-linear relationships can be accommodated. For GAMS, the usual linear function of an independent variable is replaced by an unspecified smooth function, ie. the function is suggested by the data rather than imposed upon it.

Computation and use of trigger values for physical and chemical stressors (including toxicants)

For some objectives, the risk-based framework of the *Water Quality Guidelines* for physical and chemical variables provides for an identified trigger value and an assessment of a series of measurements against this trigger value to determine if further action or monitoring is needed. An approach was developed which is outlined in Section 7.4.4 of the *Water Quality Guidelines* and detailed in Appendix 7 of Volume 2 of that document.

Trigger values are an 'early warning' mechanism to alert managers of a potential problem. **They are not intended to be an instrument to assess 'compliance' and should not be used in this capacity**. Trigger values are derived preferably from locally appropriate control or reference data, although the *Water Quality Guidelines* provide default values were such data do not exist or cannot be gathered.

In formal terms the trigger-base approach is as follows: **A trigger for further investigation will be deemed to have occurred when the median concentration of n independent samples taken at a test site exceeds the eightieth percentile of the same indicator at a suitably chosen reference site or from the relevant guideline value in the Water Quality Guidelines.**

This approach is statistically-based and acknowledges natural background variation by comparison to a reference site. It is robust in that it accommodates site-specific anomalies and uses a robust statistical measure as the basis for triggering. No assumptions are required to be made about the distributional properties of the data obtained from either the test or reference sites. The computational requirements of the approach are minimal and can be performed without the need for statistical tables, formulae, or computer software. Finally, the temporal sequence of trigger events is readily captured in a simple plot.

The procedure is responsive to shifts in the *location* (ie. 'average') of the distribution of values at the test site. While differences in shape of the reference and test distribution may be important in some instances, this is a secondary consideration that is not specifically addressed by this protocol. It is also important to note that the role of the 80^{th} percentile at the reference site is to simply quantify the notion of a 'measurable perturbation' at the test site. The protocol is not a statistical test of the equivalence of the 50^{th} and 80^{th} percentiles *per se*. The advantages of using a percentile of the reference distribution are 1) it avoids the need to specify an absolute quantity, and 2) because the reference site is being monitored over time, the trigger criterion is being constantly updated to reflect temporal trends and the effects of extraneous factors (eg. climate variability, seasonality).

Implementation of the trigger criterion is both flexible and adaptive. For example, the user can identify a level of routine sampling (through the specification of the sample size *n*) that provides an acceptable balance between cost of sampling and analysis and the risk of false triggering. The method also encourages the establishment and maintenance of long-term reference monitoring as an alternative to comparisons with the default guideline values provided in Section 3.3 of the *Water Quality Guidelines* that do not account for site-specific anomalies.

The steps in implementing this procedure are summarised below, but readers are encouraged to familiarise themselves with Appendix 7 of the *Water Quality Guidelines* where it is elaborated in detail with worked examples.

1. Data requirements for the reference condition

Prior to implementing the trigger rule, the user will need to have selected appropriate reference sites as described in Chapter 3 of the *Monitoring and Reporting Guidelines* and Section 3.1.4 of the *Water Quality Guidelines*. In addition, for physical and chemical indicators taken from the water column, the **minimum** data requirements at the reference site should consist of **two years of contiguous monthly data** before a valid trigger value can be established. Until this minimum data requirement has been established, comparison of the test site median should be made with reference to the default guideline values identified in Section 3.3 of the *Water Quality Guidelines*.

Data from toxicants will often be a special case when implementing the *Water Quality Guidelines* in that they will usually be compared with a single default trigger value which has been prepared by analysis of a comprehensive set of the available ecotoxicological data rather than comparison with data from a specific reference site. Situations where the default guideline value should be varied when deciding upon the trigger value are detailed in Section 7.4.4.2 of the *Water Quality Guidelines*.

For physical and chemical indicators from sediments, the temporal scope of sampling used to establish the reference condition will be inappropriate because accumulation rates of sediments are very slow (typically < 10 mm/yr). It is more appropriate to use spatial variability, either based on depth profiles at a test site or an appropriate number of surface sediment samples, to characterise an appropriate reference condition. A number of other issues specific to sediments (eg. influence of grain size) also need to be taken into account, and these are described in Section 7.4.4.4 of the *Water Quality Guidelines*.

2. Computation of the 80th percentile at the reference site

The computation of the 80th percentile at the reference site is always based on the *most recent* 24 monthly observations. The procedure is as follows:

(i) arrange the 24 data values in *ascending* (ie. lowest to highest) order,

(ii) take the simple average (mean) of the 19th and 20th observations in this ordered set.

3. Updating the reference site data and 80th percentile

Each month, a new reading at the reference (and test) site is obtained. The reference site observation is appended to the end of the original (ie. unsorted) time sequence. Steps (i) and (ii) from 2 above are applied to the most recent 24 data values. Note, even though only the most recent two years of data is used in the computations, no data should be discarded; maintenance of the complete data record will allow longer-term statistics to be computed.

4) Data requirements at the test site

A feature of the method is the flexibility it provides the user for the allocation of resources to the sampling effort. As previously mentioned, there is no fixed requirement to monitor at a reference location (ie. the default guideline values can be applied). Similarly, the choice of sample size at the test site is arbitrary, although there are implications for the rate of false triggering. For example, a minimum resource allocation would set n=1 for the number of samples to be collected each month from the test site. It is clear that the chance of a *single* observation from the test site exceeding the 80th percentile of a reference distribution which is *identical* to the test distribution is precisely 20%. Thus the Type I error in this case is 20%. This figure can be reduced by increasing n. For example, when n=5 the Type I error rate is approximately 0.05. The concomitant advantage of larger sample sizes is the reduction in Type II error (the probability of a false no-trigger). So-called 'power curves' are provided in Appendix 7 (Volume 2) of the *Water Quality Guidelines* to help understand the effects on error rates of a particular sampling strategy at the test site.

5) Computation of the median at the test site
The median is defined to be the 'middle' value in a set of data such that half of the observations have values numerically greater than the median and half have values numerically less than the median. For small data sets, the sample median is obtained as either the single middle value after sorting in ascending order when *n* is *odd*, or the average of the two middle observations when *n* is *even*.

6) Use of the control chart
The foregoing has been provided to assist with the month-by-month comparisons. It is suggested that these monthly results be plotted in a manner indicated in Figure 3. This provides a visual inspection of all results and helps identify trends, anomalies, periodicities and other phenomena. The methods in Chapter 6 of the *Monitoring and Reporting Guidelines* can be used to model trends and other data behaviour if required.

7) Comparing test data against single guideline (default values)
In the absence of suitable reference site data, for most physical and chemical indicators the median of the test site data is to be compared with the default guideline value from the *Water Quality Guidelines* as illustrated in Figure 4. This guideline value has been computed as the 80th percentile of the amalgamation of a number of historical data sets across broad geographical regions. Unlike the comparison with a locally-derived 80th percentile, the guideline value is static and will not reflect any local spatial and/or temporal anomalies. Reference site monitoring is strongly advocated if these effects are considered to represent a significant source of departure from the guideline value.

For toxicants a more conservative approach taken, where it is recommended that action is triggered if the 95th percentile of the test distribution exceeds the default value. This more stringent approach is warranted because, unlike other physical and chemical stressors, the default values for toxicants are based on observed biological effects implying that exceedance of the trigger value indicates the potential for ecological harm. Note that because the proportion of values required to be less than the default trigger value is very high (95%), a single observation greater than the trigger value would be legitimate grounds for action in most cases, even early in a sampling program.

Similarly, because of the poor reliability of the sediment trigger values it is difficult to be prescriptive about how these can be compared with test values. The same applies to the comparison of reference site values with test sites, where comparisons of reference median or 80th percentile with the test site median may be equally appropriate in giving an estimate of the relative concentrations, which is really all that is required in the case of sediments. However, where sediment samples within a test site clearly exceed trigger values, or are reasonably inferred to be ecologically hazardous, the *Water Quality Guidelines* recommend additional sampling to more precisely delineate contaminated zones within a site.

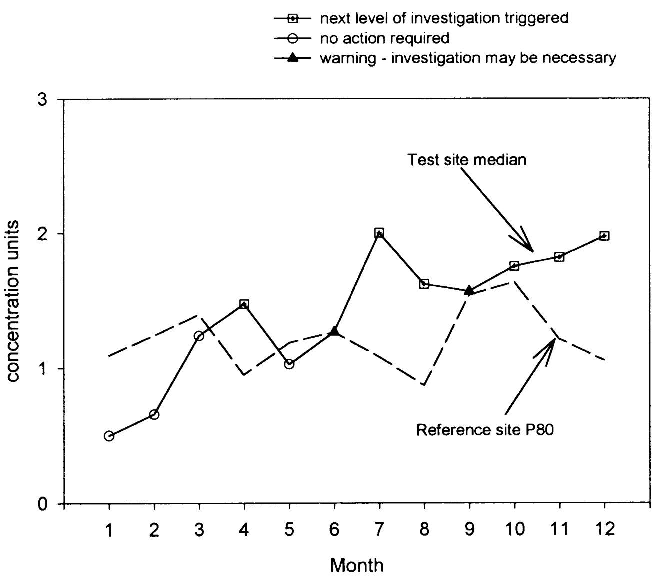

Figure 3. *Control chart showing physical and chemical data (Y axis) for test and reference sites plotted against time, and recommended actions (after ANZECC & ARMCANZ 2001a).*

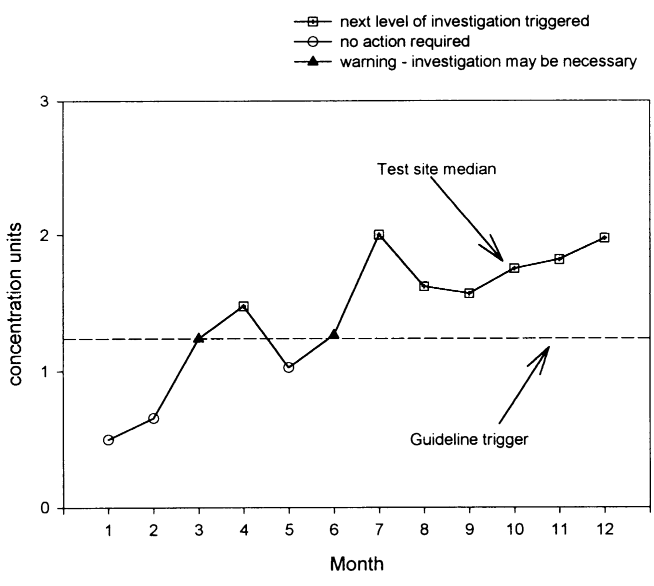

Figure 4. *Control chart showing physical and chemical data (Y axis) for test site plotted against default trigger value, time and recommended actions (after ANZECC & ARMCANZ 2001a).*

REPORTING

Reporting is not just confined to publication of results at the conclusion of the program. While designing the study, the monitoring team should identify regular reporting requirements that should occur during the course of the study which will be necessary to adjust or refine the sampling design or the field and laboratory procedures. Often there will also be regular reporting required by the agency funding the study in the form of progress or milestone reports. At the completion of the study, dissemination of the interpreted results may require using different formats and media to communicate clearly with the different stakeholders involved in the issue for which the program was developed.

Thus there are often different audiences with different levels of expertise who will be 'consumers' of the outputs from a monitoring program. The frequent, regular reporting schedules within the monitoring team will ordinarily require little interpretation of the outputs because the staff involved are already trained in the interpretation of their results and those of their QA/QC procedures. Progress and completion reports from a program should interpret the information so that the objectives and outcomes of the investigation are clearly articulated to those who have to make decisions based on the program. Therefore, a short executive summary is an important part of such reports, as well as sufficient technical detail in the body of the report to allow readers to judge the efficacy and reliability of the information on which the conclusions are based. The completion report should become the primary source on which further communication activities are based, and peer review of such reports is encouraged since this increases public confidence in the outcomes. Public comment or review of drafts of the report are also warranted in many situations for similar reasons.

Publication of findings in peer-reviewed journals and presentation of results to scientific meetings are familiar to most water quality practitioners. Communicating to other audiences, however, requires careful consideration of their requirements. A variety of media can be used (eg. newsletters, video, internet web pages), but other activities such as public meetings or open days can also be very effective under some circumstances. When publicising outcomes in the popular (eg. broadcast and newspaper) media, information should be disseminated in well-structured, authorised media releases. These media releases are best prepared by an officer with professional training in communicating information, and should include appropriate contact information. Personnel being interviewed by the print media should ask to view a transcript of the article before it is published.

ACKNOWLEDGEMENTS

We are grateful for the help and comments provided by Dr David Fox (CSIRO Land and Water) and Dr. John Chapman (NSW Environment Protection Authority).

REFERENCES

ANZECC & ARMCANZ. 2001a. *Australian and New Zealand guidelines for fresh and marine waters. National Water Quality Management Strategy Paper No 4*, Australian and New Zealand Environment and Conservation Council & Agriculture and Resource Management Council of Australia and New Zealand, Canberra, Australian Capital Territory, Australia.

ANZECC & ARMCANZ. 2001b. *Australian guidelines for water quality monitoring and reporting. National Water Quality Management Strategy Paper No 7*, Australian and New Zealand Environment and Conservation Council & Agriculture and Resource Management Council of Australia and New Zealand, Canberra, Australian Capital Territory, Australia.

Bardwell DA. 1991. Problem-framing: a perspective on environmental problem solving. *Environ. Manage.* **15,** 603-612.

Beyers DW. 1998. Causal inference in environmental impact studies. *J. North Am. Benthol. Soc.* **17,** 367-373.

Chapman PM. 1990. The Sediment Quality Triad approach to determining pollution-induced degradation. *Sci. Tot. Environ.* **97-8,** 815-825.

Cressie NAC. 1993. *Statistics for Spatial Data*, John Wiley and Sons, New York, NY, U.S.A.

Dobson AJ. 1990. *An introduction to generalised linear models*, Chapman and Hall, London.

Fox DR. 1999. Setting water quality guidelines - a statistician's perspective. *SETAC News* **May 1999,** 7-18.

Green RH. 1979. *Sampling Design and Statistical Methods for Environmental Biologists*, John Wiley and Sons, New York, NY, U.S.A.

Hastie TJ and Tibshirani RJ. 1990. *Generalized additive models*. Chapman and Hall, London.

Humphrey CH, Barmuta LA, Davies PE, Davis JA, Mapstone BD, Trayler K, Ward TJ and Biggs BJF. 2001. Incorporating biological assessment into new generation water quality guidelines in Australia and New Zealand. *Australas. J. Ecotox.* **7**, 65-80.

Humphrey CL, Faith DP and Dostine PL. 1995. Baseline requirements for assessment of mining impact using biological monitoring. *Aust. J. Ecol.* **20,** 150-166.

Johnson DH. 1999. The insignificance of statistical significance testing. *Journal of Wildlife Management* **63**, 763-772.

Maher W and Cullen P. 1997. Design of nutrient programs addressing the Why question. In *Proceedings of Workshop on Sampling Nutrients in Aquatic Environments: Collecting Valid and Representative Samples*, Markwort KN and Maher WA. (Eds), CRC for Freshwater Ecology, University of Canberra and National Eutrophication Management Program, Canberra, Australian Capital Territory, Australia, pp 2-9.

McBride GB, Loftis JC and Adkins NC. 1993. What do significance tests really tell us about the environment? *Environmental Management* **17**, 423-432.

McCullagh P and Nelder JA. 1983. *Generalized linear models*, Chapman and Hall, London.

Stewart-Oaten A and Bence JR. 2001. Temporal and spatial variation in environmental impact assessment. *Ecol. Monogr.* **71,** 305-339.

Thompson SK. 1992. *Sampling*, John Wiley and Sons, New York, NY, U.S.A.