National Water Quality Management Strategy

No. 7

AUSTRALIAN GUIDELINES FOR WATER QUALITY MONITORING AND REPORTING

October 2000

Australian and New Zealand Environment and Conservation Council

Agriculture and Resource Management Council of Australia and New Zealand

Copies of this publication may be obtained from:

Australian Water Association
PO Box 388
ARTARMON NSW 2064
Telephone (02) 9413 1288 Facsimile (02) 9413 1047

OR Australian Government Info Shops in all capital cities and Townsville. For locations, contact and ordering details check http://www.dofa.gov.au/infoaccess/general/purchase_info_products.html or phone 132 447 (toll free in Australia, 24 hour service).

OR NZ Water and Wastes Association
PO Box 13880
Onehunga
AUCKLAND 1006 New Zealand
Telephone 64 9 636 3636
Facsimile 64 9 636 1234
Email water@nzwwa.org.nz

For further information on acknowledgment, contact:

The Secretary
Agriculture and Resource Management Council of Australia and New Zealand
GPO Box 858
CANBERRA ACT 2601
Telephone (02) 6272 5216 Facsimile (02) 6272 4772

OR

The Secretary
Australian and New Zealand Environment and Conservation Council
GPO Box 787
CANBERRA ACT 2601
Telephone (02) 6274 1428 Facsimile (02) 6274 1858

Environment Australia Cataloguing-in-Publication information

Australian guidelines for water quality monitoring and reporting
/ Australian and New Zealand Environment and Conservation Council,
Agriculture and Resource Management Council of Australia and New Zealand.

(National Water Quality Management Strategy ; no.7)

Bibliography.
Includes index.

ISBN 0 642 19562 5
ISSN 1038 7072

1. Water quality – Australia – Measurement. 2. Water quality management – Australia.
I. Australian and New Zealand Environment and Conservation Council. II. Agriculture and Resource Management Council of Australia and New Zealand. III. Series

628.161'0994-dc21

Printed in Australia on recycled paper for the Australian and New Zealand Environment and Conservation Council, and the Agriculture and Resource Management Council of Australia and New Zealand

Table of Contents

Chapter Four: Field sampling program

Chapter Five: Laboratory Analysis

TABLES

BOXES

FIGURES

Do you want to comment on the Monitoring Guidelines? Use the email and postal addresses on page xii.

Preamble

The National Water Quality Management Strategy (NWQMS) is a nationally agreed set of policies, processes and 21 guidelines documents (Appendix 1). It is being developed jointly by the Agriculture and Resource Management Council of Australia and New Zealand (ARMCANZ) and the Australian and New Zealand Environment and Conservation Council (ANZECC). The NWQMS has also been endorsed by the Council of Australian Governments (COAG), which represents all three levels of government in Australia.

This document, the *Australian Guidelines for Water Quality Monitoring and Reporting*, is an integral element of the NWQMS. It provides a comprehensive framework and guidance for the monitoring and reporting of fresh and marine waters and groundwater. The document does not discuss drinking water, waste water and effluents; they are covered by separate NWQMS guidelines.

Worldwide, the quality of surface and groundwater, estuarine and marine waters tends to decline because of human activities. Concerted management and action by government, community and industry can reduce or reverse the decline in water quality, and that is the basis of Australia's NWQMS. Water quality must be measured, i.e. monitored, regularly and the results analysed, interpreted, reported and acted upon to achieve effective concerted management.

The *Australian Guidelines for Water Quality Monitoring and Reporting* provides a comprehensive framework and guidance for monitoring and reporting, but it should be used in conjunction with other NWQMS technical papers, especially the *Australian and New Zealand Guidelines for Fresh and Marine Water Quality*, paper no.4 in the NWQMS series. Standard reference works and documented in-house procedures should also be consulted and followed, subject to appropriate quality assurance and quality control criteria, and any monitoring programs resulting from the use of this guidelines document should be consistent with relevant local and state regulations and by-laws.

The *Australian Guidelines for Water Quality Monitoring and Reporting* describes the design, application, analysis and reporting of monitoring programs. Each chapter presents a flowchart and checklist of actions necessary for effective monitoring and reporting, as well as information about the various stages of the operation. Extra details are given in appendices and the whole text contributes to the index at the end of the volume. A shorter version, the *Australian Guidelines for Water Quality Monitoring and Reporting — Summary*, accompanies the main document. It comprises flowcharts and a summary of the text of the main document, but it does not include checklists: instead, there the text is arranged as a series of expanded checklists. Both versions contain a glossary of major terms and an index.

The *Australian Guidelines for Water Quality Monitoring and Reporting* has been developed with major input by environment and water agencies and other parties throughout Australia.

To be continuously relevant to its users, the *Australian Guidelines for Water Quality Monitoring and Reporting* (the Monitoring Guidelines), like other NWQMS benchmark documents, will require ongoing review and revision. The present version was current up to October 2000. Users are invited to comment on the *Australian Guidelines for Water Quality Monitoring and Reporting* by contacting the offices listed overleaf. These addresses can also receive comments on the *Australian and New Zealand Guidelines for Fresh and Marine Water Quality* (the Water Quality Guidelines), so users should name the document to which their comments apply.

Disclaimer

The contents of this document have been compiled using a range of source materials, and, while reasonable care has been taken in its compilation, the member governments of ANZECC and ARMCANZ and the organisations and individuals involved with the compilation of this document shall not be liable for any consequences which may result from using the contents of this document.

Contacts for comments and information

If you want further information or advice about this guidelines document or to comment on aspects of it, such as possible errors or omissions, or changes required for future revisions, please contact the agency designated below for *your state or territory* in Australia or for New Zealand.

Australian Capital Territory

Bob Neil
Environment ACT
PO Box 144, Lyneham, ACT 2602
Tel: (02) 6207 2581 Fax: (02) 6207 6084
email: robert.neil@act.gov.au

New South Wales

Pollution Line
NSW Environment Protection Authority
PO Box A290, Sydney South, NSW 1232
Tel: 131 555 Fax: (02) 9995 5911
email: info@epa.nsw.gov.au

Northern Territory

Director of Resource Management
Natural Resources Division
Department of Lands Planning & Environment
PO Box 30, Palmerston, NT 0831
Tel: (08) 8999 4455 Fax: (08) 8999 4403
email: michael.lawton@nt.gov.au

Queensland

Andrew Moss
Environmental Protection Agency
PO Box 155
Brisbane Albert Street, QLD 4002
Tel: (07) 3896 9245 Fax: (07) 3896 9232
email: andrew.moss@env.qld.gov.au

South Australia

John Cugley
South Australian Environment Protection Agency
GPO Box 2607, Adelaide SA 5001
Tel: (08) 8204 2055 Fax: (08) 8204 2107
email: cugley.john@saugov.sa.gov.au

Tasmania

Director of Environmental Management
Scientific and Technical Branch, Water Section
Dept of Primary Industries, Water and Environment
GPO Box 44A, Hobart, TAS 7001
Tel: (03) 6233 6518 Fax: (03) 6233 3800
email: waterinfo@dpiwe.tas.gov.au

Victoria

Lisa Dixon
Manager Freshwater Sciences
Environment Protection Authority
GPO Box 4395QQ, Melbourne, VIC 3001
Tel: (03) 9616 2361 Fax: (03) 9614 3575
email: lisa.dixon@epa.vic.gov.au

Western Australia

Victor Talbot
Department of Environmental Protection
PO Box K822, Perth
WA 6842
Tel: (08) 9222 8655 Fax: (08) 9322 1598
email: victor_talbot@environ.wa.gov.au

New Zealand

Nigel Bradly
Land and Water Group
Ministry for the Environment
PO Box 10362, Wellington NEW ZEALAND
Tel: NZ (04) 917 7489 Fax: NZ (04) 917 7523
Mobile: NZ 025 379 391
email: nigel.bradly@mfe.govt.nz

Acknowledgments

The *Australian Guidelines for Water Quality Monitoring and Reporting* has been developed in several stages. The completed document largely derives from a consultancy (see (i) below) to restructure and expand an earlier draft (see (iii) below) of the document. More recently, two editing consultancies (see (ii) below) have provided major input, including text for Chapter 6 and much of Appendix 5.

(i) Developing the *Australian Guidelines for Water Quality Monitoring and Reporting* — Expansion of an Earlier Draft

A peer review of the earlier draft (iii) had indicated the need for major restructuring and augmentation of the document, focusing on monitoring program design and related issues. This revision was undertaken by the Cooperative Research Centre for Freshwater Ecology (CRCFE) at the University of Canberra (UC). The consultancy was funded by Environment Australia (EA) and Agriculture, Fisheries and Forestry Australia (AFFA). The persons who undertook the work were:

Bill Maher (CRCFE project leader), CRCFE,
Professor Peter Cullen, CRCFE,
Ian Lawrence, CRCFE,
Anthony Chariton, Applied Ecology, UC,
Daniel Spooner, Applied Ecology, UC,
David Judge, Applied Ecology, UC,
Peter Liston, Planning and Land Management, Department of Urban Services, ACT.

The EA Project Officer was Charles Lewis.

(ii) Editing and Finalising the *Australian Guidelines for Water Quality Monitoring and Reporting*

Following the expansion consultancy, two consultancies were let to CSIRO Environmental Projects Office to augment, edit and finalise the whole document. These major consultancies were funded by EA. The persons who undertook the work were:

Graeme Batley (lead reviewer), CSIRO,
David Fox (statistical aspects), CSIRO,
Jenny Stauber (biological monitoring), CSIRO,
Bill Maher (reviewer), CRCFE,
Ann Milligan (editing), Science Text Processors Canberra.

Robert Molloy of CSIRO Environmental Projects Office was the CSIRO Project Coordinator.

The EA Project Officer was Charles Lewis. Sharon Rees was also a project officer for the letting of the consultancy. John Anderson (EA) also worked as a project officer from March to June 2000.

Transfer of parts of the monitoring component of the draft *Australian and New Zealand Guidelines for Fresh and Marine Water Quality* to the draft *Australian Guidelines for Water Quality Monitoring and Reporting*, with rationalising, was undertaken by Leon Barmuta, Department of Zoology, University of Tasmania, in conjunction with Graeme Batley and David Fox, CSIRO, Bill Maher, CRCFE, Chris Humphrey, Environmental Research Institute of the Supervising Scientist, ***eriss***, and Charles Lewis, EA. The transferred material was initially prepared by Chris leGras, ***eriss***.

Rob Donohue, Waters and Rivers Commission, Western Australia, provided input to the discussion on monitoring in relation to water quality guidelines.

(iii) Earlier Development of the *Australian Guidelines for Water Quality Monitoring and Reporting*

The earlier development of the draft *Australian Guidelines for Water Quality Monitoring and Reporting* was based on the *Water Quality Investigations Manual: Preferred Methods for Sampling and Analysis* (1995) compiled by the NSW Environment Protection Authority (EPA) and based on a consultancy by ICF Pty Ltd, with principal consultants David Garman and Turlough Guerin. The NSW Manual was edited by Rob Mann of the NSW EPA, with significant contribution from other NSW EPA staff, especially Evelyn Goodwin, and from the EPA Chemical Laboratory.

Development of the *Australian Guidelines for Water Quality Monitoring and Reporting* was initially undertaken by a Working Group drawn from Commonwealth and state environment protection and water resource management agencies. The Working Group comprised: Lionel Wood (AFFA) (Chair), Charles Lewis (EA), Graham Rooney (Melbourne Water), Ross Higginson (NSW EPA) and Voytek Poplawski (Department of Natural Resources, Queensland).

Lance Woods was engaged by EA and AFFA to prepare additional material and edit the document that then went to a peer review. Comments received were incorporated into the document by Mary Taylor, Patty Please, Kwame Asumadu and Nick Ladner (AFFA).

For all stages of the development of the *Australian Guidelines for Water Quality Monitoring and Reporting*, comments, contributions and guidance were received from members of the ARMCANZ and ANZECC Contact Group, which oversees the development of the NQWMS. Most environment and water agencies in Australia have significantly contributed to the development of the document through comments on drafts, but special mention is made of the extensive and valuable comments provided by NSW agencies. A number of other organisations also commented on drafts of the document.

David Lambert, Stephen Clark and Rachael Gregson provided support from AFFA for the development and operation of the expansion and editing consultancies.

As the project neared completion, Richard Norris (CRCFE) and Peter Liston checked details of many references. In a consultancy for the final editing and design phase, funded by EA, Eileen Dallwitz, TechType, improved the appearance and electronic characteristics of the diagrams and the worked examples; Vern Verass, Design Direction, guided the page design of the final document; Ann Milligan, Science Text Processors Canberra, was responsible for the resulting page layout, and the index.

Clare Nolan, Clockwork Communicators, designed the front covers and spine.

Cover photographs

The photographs on the front cover, in the top row, left–right, were supplied by:
CRC Freshwater Ecology (CRCFE); Bureau of Rural Sciences (BRS); BRS; CRCFE;
and in the bottom row, left–right, by:
BRS, Great Barrier Reef Marine Park Authority (GBRMPA); CRCFE; GBRMPA.

Glossary of terms and acronyms

Acidity
the quantitative capacity of a water to react with a strong base to a designated pH

Acute toxicity
rapid adverse effect (e.g. death) caused by a substance in a living organism. Can be used to define either the exposure or the response to an exposure (effect).

AFFA
Agriculture, Fisheries and Forestry Australia

Algae
comparatively simple chlorophyll-bearing plants, most of which are aquatic, and microscopic in size

Alkalinity
the acid-neutralising capacity of an aqueous system; the sum of all titratable bases

Ambient
surrounding

ANCA
Australian Nature Conservation Agency

Anion
a negatively-charged ion

ANZECC
Australian and New Zealand Environment and Conservation Council

Aquatic ecosystem
any water environment from small to large, from pond to ocean, in which plants and animals interact with the chemical and physical features of the environment

Aquifer
an underground layer of permeable rock, sand or gravel that carries water, allowing it free passage through pore spaces

ARMCANZ
Agriculture and Resource Management Council of Australia and New Zealand

AUSRIVAS
Australian River Assessment Scheme

BACI
Before–after, control–impact

Benchmark
a standard or point of reference

Benthic
referring to organisms living in or on the sediments of aquatic habitats

Bioaccumulation
a general term describing a process by which chemical substances are accumulated by aquatic organisms from water directly or through consumption of food containing the chemicals

Bioassay

a test used to evaluate the relative potency of a chemical by measuring its effect on a living organism relative to a control

Bioavailable

able to be taken up by organisms

Biochemical oxygen demand (BOD)

the decrease in oxygen content in a sample of water that is brought about by the bacterial breakdown of organic matter in the water

Bioconcentration

a process by which there is a net accumulation of a chemical directly from water into aquatic organisms, resulting from simultaneous uptake (e.g. by gill or epithelial tissue) and elimination

Biomagnification

the result of the processes of bioconcentration and bioaccumulation by which tissue concentrations of bioaccumulated chemicals increase as the chemical passes up through two or more trophic levels. The term implies an efficient transfer of chemicals from food to consumer so that the residue concentrations increase systematically from one trophic level to the next.

Bivalve

mollusc with a shell in two parts, hinged together

Bloom

an unusually large number of organisms of one or a few species, usually algae, per unit of water

BOD

Biochemical oxygen demand or biological oxygen demand

BOD test

an empirical test that measures the relative oxygen requirements of waste-waters, effluents and contaminated waters by incubating samples in the dark at a certain temperature for a fixed number of days usually designated by a subscript, e.g. BOD_5 test

Cation

a positively-charged ion

Chemical oxygen demand (COD)

the oxygen equivalent of the organic matter content of a sample that is susceptible to oxidation by a strong chemical oxidant

Chronic toxicity

toxicity that acts over a long period of time and that typically affects a life stage (e.g. reproductive capacity); it can also refer to toxicity resulting from a long-term exposure

COAG

Council of Australian Governments

COD

Chemical oxygen demand

Community

assemblage of organisms characterised by a distinctive combination of species occupying a common environment and interacting with one another

Community composition

all the types of taxa present in a community

Concentration

the quantifiable amount of a substance in water, food or sediment

Contaminants

biological or chemical substances or entities, not normally present in a system, capable of producing an adverse effect in a biological system, seriously injuring structure or function

Control

part of an experimental procedure that is ideally exactly like the treated part except that it is not subject to the test conditions. It is used as a standard of comparison, to check that the outcome of the experiment is a reflection of the test conditions and not of some unknown general factor.

CRCFE

Cooperative Research Centre for Freshwater Ecology

Criteria (water quality)

scientific data evaluated to derive the recommended quality of water for different uses

CSIRO

Commonwealth Scientific and Industrial Research Organisation

DEST

Department of Environment, Sport and Territories

Detection limit

method detection limit is the concentration of a substance that when processed through the complete analytical method produces a signal that has a 99% probability of being different from the blank

DO

Dissolved oxygen

DOC

Dissolved organic carbon

Duplicate samples

obtained by dividing a sample into two or more subsamples, to reveal the sizes of random and/or systematic errors due to contamination

EA

Environment Australia

EC

Electrical conductivity

Ecologically sustainable development

development that improves the total quality of life, both now and in the future, in a way that maintains the ecological processes on which life depends

Electrical conductivity

the ability of water or soil solution to conduct an electric current; commonly used as a measure of salinity or total dissolved salts

Environmental values

particular values or uses of the environment that are important for a healthy ecosystem or for public benefit, welfare, safety or health and that require protection from the effects of contaminants, waste discharges and deposits. Several environmental values may be designated for a specific waterbody.

EPA

Environment Protection Authority

Epilimnion

the uppermost layer of water in a lake, characterised by an essentially uniform temperature that is generally warmer than elsewhere in the lake, and by relatively uniform mixing by wind and wave action

eriss

Environmental Research Institute of the Supervising Scientist

ESD

Ecologically sustainable development

Euphotic

surface waters to a depth of approximately 80–100 m; the lit region that extends virtually from the water surface to the level at which photosynthesis fails to occur because of reduced light penetration

Eutrophication

enrichment of waters with nutrients, primarily phosphorus, causing abundant aquatic plant growth and often leading to seasonal deficiencies in dissolved oxygen

Fate

disposition of a material in various environmental compartments (e.g. soil or sediment, water, air, biota) after transport, transformation and degradation

FNARH

First National Assessment of River Health

Guideline

numerical concentration limit or narrative statement recommended to support and maintain a designated water use

Guideline trigger levels

the concentrations (or loads) for each water quality parameter, below which there exists a low risk that adverse biological (or ecological) effects will occur. They are the levels that trigger some action, either continued monitoring in the case of low risk situations or further ecosystem-specific investigations in the case of high risk situations.

Hardness

a measure of the sum of the concentrations of calcium and magnesium ions in water, both expressed as mg/L calcium carbonate equivalent

Humic substances

heterogeneous yellow-black organic materials that include most of the naturally dissolved organic matter in water. They are classified as huin (not soluble at any pH), humic acid (not soluble at pH <2) and fulvic acid (soluble at all pH values).

Hydrograph

graphical representation of either surface stream discharges or water-level fluctuations in wells

Hypolimnion

the region of a waterbody that extends from below the thermocline to the bottom of a lake; it is thus removed from much of the surface influence, and is usually cold and relatively undisturbed

Hypothesis

supposition drawn from known facts, made as a starting point for further investigation

Index (indices)

composite value(s) that can give a quick ranking to a waterbody or other ecosystem feature, derived via a formula that combines measurements of important ecosystem characteristics; typically used to rank 'health' or naturalness

Indicator

measurement parameter or combination of parameters that can be used to assess the quality of water

Invertebrates

animals lacking a dorsal column of vertebrae or a notochord

Ion

an electrically charged atom

Leaching

the dissolution of a material, by water or another solvent mixing with a solid phase, and its downward or outward movement from the solid in solution

ICM

Integrated catchment management

Level of protection

the acceptable level of change from a defined reference condition

Management goals

long-term management objectives that can be used to assess whether the corresponding environmental value is being maintained. They should reflect the desired levels of protection for the aquatic system and any relevant environmental problems.

MBACI

Multiple before–after, control–impact

MBACIP

Multiple before–after, control–impact, paired

MDBC

Murray-Darling Basin Commission

Measurement parameter

any parameter or variable that is measured to find something out about an ecoystem

Methylation

the introduction of methyl (CH_3) groups into organic and inorganic compounds

NATA

National Association of Testing Authorities of Australia

NHMRC

National Health and Medical Research Council

Not detectable

below the limit of detection of a specified method of analysis

NRC

National Research Council

NRHP

National River Health Program

NWQMS

National Water Quality Management Strategy

OH&S

occupational health & safety

Organism

any living animal or plant; anything capable of carrying on life processes

Oxidation

the combination of oxygen with a substance, or the removal of hydrogen from it, or, more generally, any reaction in which an atom loses electrons

PAHs

polycyclic aromatic hydrocarbons

Parameter

a measurable or quantifiable characteristic or feature

Pathogen

an organism capable of eliciting disease symptoms in another organism

Pelagic

term applied to organisms of the plankton and nekton which inhabit the open water of a sea or lake

Performance indicators

indicators used to assess the risk that a particular issue will occur (they are used in the guideline packages to compare against the trigger levels). They are generally median (or mean) concentrations in the ambient water, and may be stressor and/or condition indicators.

Periphyton

organisms attached to submerged plants

Pesticide

substance or mixture of substances used to kill unwanted species of plants or animals

pH

the intensity of the acidic or basic character of a solution, defined as the negative logarithm of the hydrogen ion concentration of a solution

Plankton

plants (phytoplankton) and animals (zooplankton), usually microscopic, floating in aquatic systems

Precipitation

the settling out of water from cloud, in the form of rain, hail, fog, snow, etc. (also the formation and settling out of solid particles in solution)

Producers

organisms that can build up their body substance from inorganic materials

Protocol

a formally agreed method and procedure for measuring an indicator, including sampling, sample handling procedures and sample analysis

Pseudoreplication

replication in which the samples are not independent but instead are from sub-populations of a population: replicates that are actually subsamples of one sample are pseudoreplicates; and samples from various sites along a stretch of river are pseudoreplicates because the water is the same, moving between sites

QA/QC

quality assurance/quality control

Quality assurance (QA)

the implementation of checks on the success of quality control (e.g. replicate samples, analysis of samples of known concentration)

Quality control (QC)

the implementation of procedures to maximise the integrity of monitoring data (e.g. cleaning procedures, contamination avoidance, sample preservation methods)

Redox

simultaneous (chemical) reduction and oxidation: reduction is the transfer of electons to an atom or molecule; oxidation is the removal of electrons from an atom or molecule

Reference condition

an environmental quality or condition that is defined from as many similar systems as possible and used as a benchmark for determining the environmental quality or condition to be achieved and/or maintained in a particular system of equivalent type

Risk

a statistical concept defined as the expected frequency or probability of undesirable effects resulting from a specified exposure to known or potential environmental concentrations of a material, organism or

condition. A material is considered safe if the risks associated with its exposure are judged to be acceptable. Estimates of risk may be expressed in absolute or relative terms. Absolute risk is the excess risk due to exposure. Relative risk is the ratio of the risk in the exposed population to the risk in the unexposed population.

RIVPACS

River Invertebrate Prediction and Classification System

SAA

Standards Association of Australia

Salinity

the presence of soluble salts in water or soils

Sediment

unconsolidated mineral and organic particulate material that has settled to the bottom of aquatic environments

SOE

State of the Environment

Solution concentration

concentration of contaminants in the liquid phase

Speciation

measurement of different chemical forms or species of an element in a solution or solid

Species

generally regarded as a group of organisms that resemble each other to a greater degree than members of other groups and that form a reproductively isolated group that will not normally breed with members of another group. (Chemical species are differing compounds of an element.)

Species richness

the number of species present (generally applied to a sample or community)

Stakeholder

a person or group (e.g. an industry, a government jurisdiction, a community group, the public, etc.) that has an interest or concern in something

Standard, e.g. water quality standard

an objective that is recognised in environmental control laws enforceable by a level of government

Stressors

the physical, chemical or biological factors that can cause an adverse effect on an aquatic ecosystem as measured by the condition indicators

Sub-lethal

involving a stimulus below the level that causes death

Suspension

very small particles (solid, semi-solid, or liquid) more or less uniformly dispersed in a liquid or gaseous medium

Taxon (taxa)

any group of organisms considered to be sufficiently distinct from other such groups to be treated as a separate unit (e.g. species, genera, families)

Taxa richness

number of taxa present

TCM

Total catchment management

TDS

Total dissolved solids

Thermocline

a region or layer of water in a lake, between the well-mixed surface layer and the cold still bottom layer, where the temperature changes rapidly with respect to depth

TIE

Toxicity identification and evaluation

Toxicant

a chemical capable of producing an adverse response (effect) in a biological system, seriously injuring structure or function or producing death. Examples include pesticides, heavy metals and biotoxins (i.e. domoic acid, ciguatoxin and saxitoxins).

Toxicity

the inherent potential or capacity of a material to cause adverse effects in a living organism

Toxicity test

the means by which the toxicity of a chemical or other test material is determined. A toxicity test is used to measure the degree of response produced by exposure to a specific level of stimulus (or concentration of chemical).

Trophic level

a notional stage in the 'food chain' that transfers matter and energy through a community; primary producers, herbivores, carnivores and decomposers each occupy a different trophic level

True colour

the colour of water resulting from substances that are totally in solution; not to be mistaken for apparent colour which includes the effect of colloidal or suspended matter

UC

University of Canberra

Uptake

a process by which materials are absorbed and incorporated into a living organism

USEPA

United States Environmental Protection Agency

UWRAA

Urban Water Research Association of Australia

WHO

World Health Organization

WMO

World Meteorological Organization

Zooplankton

see plankton

Executive Summary

The *Australian Guidelines for Water Quality Monitoring and Reporting* has been developed as a benchmark document of the National Water Quality Management Strategy. It relates closely to the revised *Australian and New Zealand Guidelines for Fresh and Marine Water Quality* (ANZECC & ARMCANZ 2000), known as the Water Quality Guidelines, and provides the guidance necessary for designing monitoring programs with which to assess water quality in freshwater, marine waters and groundwaters.

The *Australian Guidelines for Water Quality Monitoring and Reporting* (the Monitoring Guidelines) encourages a national, coordinated, efficient, quality-assured and consistent approach to water quality monitoring. The resulting data should be able to be compared across regions and through time, and integrated or collated to report trends, and the resulting information should be able to be trusted to form a sound basis from which to manage water quality and the Australian environment.

Monitoring consists of a systematic and planned series of measurements or observations that are appropriately analysed and reported, with the aim of providing information and knowledge about a water body. Monitoring (and reporting) of water quality is important for environmental protection policies and programs, and for managing water resources and controlling contaminants. It underpins State of the Environment reporting, and National Audit reporting. The information that water quality monitoring generates not only describes changes in water quality but also helps explain how ecosystems function.

The Monitoring Guidelines sets out an overall framework that embraces all aspects of a monitoring program. It discusses (and illustrates by case studies) methods and routines for, and mental approaches to, the setting of monitoring program objectives, study design, field sampling, laboratory analyses, data analysis, and reporting and information dissemination. The chosen objectives of a monitoring program should answer carefully defined questions about specific issues. The objectives should be chosen after a conceptual model has been agreed upon: that is, a description of the system as it is understood at that time, based on available information and discussion with all stakeholders. The validity of the model should be tested through a number of hypotheses that form the basis of the detailed study design. The model should be continually refined and the hypotheses should be continually restated as the information collected by the study is evaluated.

Some studies are designed to be descriptive, gathering and analysing data to document the state of a system. Other studies are designed to increase our understanding of a system or to measure its reaction to change. In the latter case, appropriate statistical designs and ways of assessing cause and effect must be decided upon. The spatial boundaries, the scale of the study and its duration are fundamental, as are the detailed issues: for example, site selection, finer spatial and temporal considerations, and the required number of samples to detect change. Measurement parameters must be agreed upon, for assessing the system's physical, chemical and biological characteristics consistent with the new Water Quality Guidelines approach to integrated monitoring. Biological assessment includes ecotoxicological methods, and ecological methods such as AUSRIVAS, based on biotic indices.

Before the start of field sampling programs, the most suitable sampling methods and sampling equipment must be chosen, for surface waters, groundwaters, sediments or aquatic biota. Suitable types of containers must be selected and prepared, together with methods for sample preparation and preservation. Quality assurance and quality control (QA/QC) considerations, and issues relating to occupational health and safety should be given considerable attention, as should cost-effectiveness and the practical aspects of sampling from a wide range of water and flow conditions.

The details of laboratory analysis must also be planned; the substances or organisms to be analysed or counted determine the methods and laboratory equipment used, and the cost. Here QA/QC considerations are especially important.

The analysis of monitoring data can require expertise in statistics. The approach taken by the Monitoring Guidelines is to provide adequate information to guide a technically qualified person in analysing data. For assessing water quality in comparison with a guideline, or evaluating trends, or establishing cause and effect, differing approaches and analytical tools are needed.

Finally, the information that the monitoring program has generated must be reported. This can be done in a variety of ways, suited to various audiences, starting from a primary report in an accepted format, leading to less technical vehicles that address all possible user groups. Reporting can involve Internet web pages, media releases, newsletters or other scientific or industry publications, as appropriate.

The Monitoring Guidelines has been prepared in consultation with Commonwealth, state and territory environment and water agencies. It should provide valuable and unified guidance to the many bodies involved in, or planning, water quality monitoring throughout Australia.

Chapter One
Introduction

1.1. Background

All the governments and Ministerial Councils of Australia agree that Australia's water resources need ecologically sustainable management; that is, management with a long-term perspective, for this and future generations. Ecologically sustainable management will ensure that:

- high quality water is available for consumption;
- adequate water supplies are available for both agricultural and industrial production;
- ecological values are enhanced and protected; and
- the community's need for water-based recreation and related amenities are met.

Therefore, the Australian and New Zealand Environment and Conservation Council (ANZECC) and the Agriculture and Resource Management Council of Australia and New Zealand (ARMCANZ) have formulated the National Water Quality Management Strategy (NWQMS) (see Appendix 1 for details), with the objective of achieving

> sustainable use of the nation's water resources by protecting and enhancing their quality while maintaining economic and social development.

The National Health and Medical Research Council (NHMRC) is involved in aspects of the NWQMS that affect public health. The NWQMS has been adopted as part of the Council of Australian Governments (COAG) Water Reform Framework (Appendix 2).

ANZECC and ARMCANZ are pursuing the NWQMS policy objective by organising the production of high-status national guidelines which are to be implemented locally. The *Implementation Guidelines* (ANZECC & ARMCANZ 1998), NWQMS paper no. 3, outlines the process for protecting water quality. It recommends the use of both regulatory and market-based approaches, in a catchment management framework with community involvement. The *Australian and New Zealand Guidelines for Fresh and Marine Water Quality* (ANZECC & ARMCANZ 2000), NWQMS paper no. 4, provides guidelines for water quality and is complemented by the *Australian Guidelines for Water Quality Monitoring and Reporting*, NWQMS paper no. 7, 'the Monitoring Guidelines', this document. Monitoring is an integral part of the NWQMS, and the Monitoring Guidelines aims to ensure that water quality monitoring in Australia is done in a nationally consistent, systematic and scientifically appropriate manner. This set of guidelines leads the way to better planned water quality monitoring and reporting programs which should collect good quality data and provide useful information to managers of natural resources.

The National Land and Water Resources Audit of water monitoring in Australia (EA/NLWRA 2000), indicates that state and territory agencies are engaged in 70% of water quality monitoring programs, and local government in 19%, and that Commonwealth agencies, industry, tertiary institutions and community groups do the rest. Community groups contribute to monitoring programs either individually or through structures such as integrated catchment management committees and Waterwatch and relevant state programs (Appendix 3). With such a wide range of organisations providing data, incompatibilities can occur, sometimes making it difficult to interpret and compare results.

1.2. Scope of Water Quality Monitoring

Australia is a continent encompassing many diverse water bodies. The water resources of greatest significance include drinking water supplies and catchments, rivers, lakes and aquifers, and estuarine and marine waters that have economic, ecological or cultural value. There is no need to monitor them all; nor are there enough personnel and money for the monitoring of all the resources that need it. Risk assessment of the potential impacts of declining water quality can be used to identify the water resources or water bodies that should be monitored first.

Water quality investigations are undertaken to provide information on the health of water bodies and for the management of catchments and water resources and the environment. They may be single studies to examine a particular issue, or they may be ongoing monitoring programs. Monitoring programs can assess water quality in comparison with water quality objectives, and consider measurement parameters, ranges of concentrations and frequency of measurement, and identify point and non-point sources of contaminants. For example, a monitoring program might be able to demonstrate a relationship between changes in land management and the frequency of algal blooms in a catchment. A uniform, coordinated and efficient series of monitoring programs will improve the information available: for example,

- to provide a basis for auditing contaminant controls and assessing impacts on water quality;
- to underpin the State of the Environment reporting and National Audit reporting;
- to develop better water quality standards and guidelines and to assess water quality against these.

Water quality investigations are expensive, and few organisations have the resources to monitor over a large geographical area or over a long time frame. Resources tend to be targeted to meet specific regional needs. State government agencies will continue to bear the prime responsibility for setting priorities for monitoring and reporting within their own jurisdictions and for meeting nationally agreed objectives. As these agencies improve the coordination of monitoring they will eliminate duplication and gaps in information collection, particularly if they also draw on supplementary information collected by initiatives such as Landcare, State of the Rivers, or State of the Environment reporting (Environment Australia 1996, 1998), as appropriate.

It can be difficult to aggregate data to provide statewide or national reports, such as for State of the Environment reporting. The difficulties are reduced when standard approaches are used for the design, implementation and reporting of water quality monitoring programs, at least with respect to key measurement parameters. Ideally, the design of each program ensures that the data are collected or generated in a form that can be integrated and compared with similar data collected elsewhere in Australia; for example, the program can measure a set of parameters at the same time-intervals as have been used in a comparable program elsewhere. Reports on the monitoring program should include information about the collection, management, analysis and storage techniques used. Communication between water protection groups about a monitoring program that is being planned can also help make concurrent programs compatible with each other. Data sets can be collated and compared at a later stage to develop a wider picture from the data.

1.3. Structure of a Monitoring Program

Effective water quality investigations systematically collect physical, chemical and biological information, and analyse, interpret and report those measurements, all according to a carefully pre-planned design which follows a basic structure.

This Monitoring Guidelines document sets out a standard structure for the design of a monitoring program (see Figure 1.1). The chapters lead the monitoring team through the necessary stages. Each chapter contains a summary flowchart and checklist, and discusses how to:

- define information requirements and objectives for monitoring programs (Chapter 2);

- design a study, including its type, scale, measurement parameters and sampling programs, and preferred methods for sampling (Chapters 3 and 4);
- design a laboratory program including preferred methods for laboratory and field analysis (Chapters 4 and 5);
- set up quality assurance and quality control procedures (Chapters 4 and 5);
- be aware of occupational health and safety concerns (Chapters 4 and 5);
- statistically analyse and interpret the data (Chapter 6 and Appendix 5);
- report and disseminate information to various audiences, and collate feedback (Chapter 7).

Sometimes, more detailed advice will be required and this can either be found in the appendixes or in references or other listed sources.

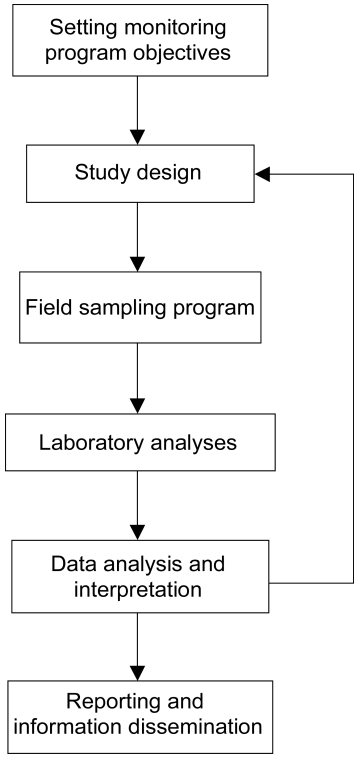


Figure 1.1. Framework for a water quality monitoring program.
Each box is dealt with by one chapter of the Monitoring Guidelines,
from Chapter 2 to Chapter 7.

It is important to remember that the design of a monitoring program is an iterative process, as indicated in Figure 1.1, and that earlier components in the structure should be refined on the basis of findings in later stages.

The Monitoring Guidelines is intended for use by water quality personnel with basic technical training, involved in environmental monitoring throughout Australia, working in agencies, water authorities, catchment management authorities, councils, industry, consulting companies and tertiary institutions, and it should also be helpful for community groups.

Chapter Two
Setting Monitoring Program Objectives

2.1. Introductory Overview

When a monitoring team plans a water quality investigation or monitoring program, it must state its objectives clearly; otherwise it will not be able to fully address the more detailed questions of how to undertake the investigation. The objective of an effective monitoring program is to provide information and knowledge about an issue, preferably for the least cost, to inform those who have commissioned and will use the data. Good monitoring programs are not just exercises in data collection.

Before defining the objectives and information requirements, the first step is to identify the issues that are to be addressed. After a comprehensive analysis of the issues, the monitoring team should understand what information is needed, and be able to formulate the specific objectives for the monitoring program.

Water quality management issues in Australia typically fall into four categories:

- the long-term management, protection and restoration of aquatic ecosystems so they can fulfil their environmental values;
- contaminants, their sources and fates in aquatic ecosystems, the magnitude of the problem and the actions that need to be taken to protect the environmental values;
- the performance of management strategies;
- conformity with water quality guidelines.

These sorts of issues have driven many monitoring programs in the past. Many monitoring programs have set out to collect information relevant to the environmental values (formerly called 'beneficial uses') of a water body. Environmental values reflect the uses that can be made of the water body, perhaps by aquatic ecosystems, or as water supply for primary industries (irrigation, stock drinking water, agriculture and aquaculture), or for recreational use and aesthetics, or for drinking water. The *Australian and New Zealand Guidelines for Fresh and Marine Water Quality*, from now on termed the Water Quality Guidelines (ANZECC & ARMCANZ 2000) has been developed so that these values can be protected.

Monitoring of waters is commonly undertaken to meet one of the following general objectives:

- to measure the quality of ambient freshwater or marine water;
- to provide assurance that the water meets appropriate guidelines for its designated use;
- to investigate why the water may not be meeting such guidelines;
- to assess the loads of materials entering the water body from the catchment (export studies);
- to assess the loads of materials carried past various points, the transformations of materials and the rates of loss in-stream or over-bank, so that streamflow mass balances can be calculated;
- to characterise the biota within a river, estuary or coastal marine water body;
- to assess biological productivity;
- to assess the state of the resource as defined by a variety of measurement parameters or indicators (State of the Environment reporting, and National Audit reporting);

- to assess the effectiveness of actions for contaminant control, or restoration or rehabilitation of waters;
- to identify trends in the condition of the water body.

This chapter outlines the process for translating issues into monitoring program objectives, as illustrated in Figure 2.1 and Table 2.1.

As part of the objective-setting exercise, it is instructive to make a preliminary assessment of the issue and then develop a conceptual model that can form the basis of the proposed monitoring study.

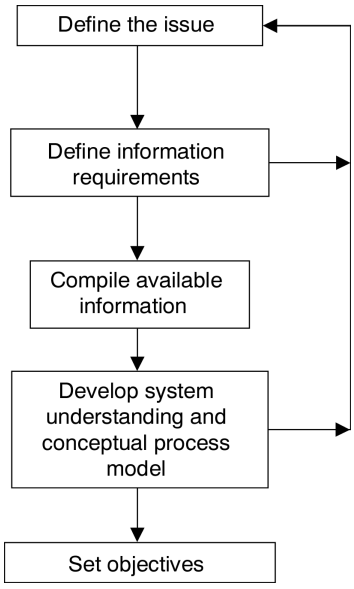


Figure 2.1. Framework for setting monitoring program objectives

Table 2.1. Checklist for determining information needs and monitoring program objectives

1.	Has the issue or question been defined?
2.	Have the identities of all the information users been ascertained, so that information is obtained that will address all the stakeholders' needs?
3.	Has all the available information relating to the issue or problem been collected, checked, and put into a common form?
4.	Have knowledge gaps been identified and the information obtained, or have the limitations and restrictions of not having that information been evaluated?
5.	Has a shared conceptual process model of the system been developed and made explicit?
6.	Have the assumptions underlying the model been made explicit?
7.	Has an analysis been undertaken to identify the essential information required?
8.	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?

2.2. Defining the Issue

To define the issue, problem or question to be answered by the monitoring program, its monitoring team must interact with the end-users of the information and the stakeholders for that area. The stakeholders may be individual residents, a community group, an industry group, a government jurisdiction, and they may be found in the local area or downstream or upstream.

The definition of the issue or problem will emerge during or after discussions between the stakeholders and the monitoring team. It will be a result of the values they hold to be important, their previous knowledge and their experience. The manner in which the problem is viewed may be a major factor in determining its outcome (Miller et al. 1960), but the initial statement of the problem or question may be the most crucial single factor in determining whether a solution can be found. Bardwell (1991) identified some pitfalls to be avoided when specifying a problem:

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

If a problem can be redefined or reframed, and conceptually explored, the monitoring team may see a larger range of alternatives and solutions to be examined or information to be obtained, and the ultimate monitoring program may benefit.

Some typical issues for new monitoring programs could include:

- excess nutrients, leading to algal blooms;
- salinity, leading to water being unacceptable for drinking or agricultural use, and having effects on aquatic ecology;
- contaminants, having acute or chronic effects on aquatic organisms or limiting water use;
- contaminants, being accumulated by biota with potential later effects on the health of human consumers;
- microbial contamination, from human or animal wastes, making water unsuitable for drinking or recreational use;
- maintenance of dissolved oxygen;
- effects of suspended particulate matter;
- effects of temperature changes;
- effects of pH changes.

Some of these are discussed in more detail in the Water Quality Guidelines (ANZECC & ARMCANZ 2000).

2.3. Compilation of Available Information

The next step in this preliminary stage of designing a monitoring program is to collect the available information relating to the issue at hand. Depending on the issue this step could entail a comprehensive literature review of current international understanding, or a review of relevant previous monitoring information collected either for the site of interest or for other locations, or interviews and recording of observations and evidence gathered by members of the local community. It is important that scarce funds are not spent merely to repeat studies on the issue or at the site of interest. However, information gained in previous investigations will help refine the information requirements and objectives of the present monitoring program.

The monitoring team will need to identify gaps in the assembled knowledge, and fill them if possible. If they cannot find the information, they must assess the limitations and restrictions caused by not having that information.

Existing data will probably consist of water quality measurements, stream-flow records and some biological data. Some of these data may have been published; others may be in the records of various agencies or research providers. They will need collection, checking and standardising into a common form using suitable data storage practices (see section 5.4.1).

2.4. Understanding the System and Formulating Conceptual Process Models

Once the issue for monitoring has been defined and the available information about it has been assembled, it is time to decide upon the questions that the monitoring program must tackle — its objectives. This is only really possible if the monitoring team has some preliminary understanding of the ecosystem for which the monitoring program is being designed. That understanding can initially be derived from the information they have just collected, and it is best formalised in a conceptual process model of the system being examined. The model need only be a simple box diagram that illustrates the components and linkages in the system to be monitored. It presents the factors that are perceived to be driving the changes in the system and the consequences of changes to these factors. For instance, in eutrophication studies, nutrients are commonly shown as the driving factors, while chlorophyll or algal cells are the consequences. Examples of conceptual models are shown in Figures 2.2–2.5.

Conceptual process models are important in defining the 'why' questions. After they have been shared with colleagues and argued about, conceptual process models set out the monitoring team's collective knowledge, experience and perspectives of the ecosystem that is the basis of the study. The final model illustrates the team's assumptions about how the system functions and what it believes to be the important or dominant processes. It is desirable for all team members to develop their own concepts of the system, and then to discuss and integrate these conceptual models. It should not be left to one team member, however experienced, because the differences between individual models can be important in clarifying the real issues and questions and in setting objectives.

Often the conceptual model will be based on accumulated wisdom as opposed to hard data. The monitoring team needs to articulate the assumptions underlying the model and to identify the gaps in information supporting these assumptions. The assumptions need to be critically reviewed because incorrect assumptions may lead to incorrect conclusions being drawn about information needs. One objective of the monitoring program will then be to collect data to validate these assumptions. 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 or question being investigated.

During the formulation of a model, several decisions must be made or the model will be too complex:

- what is the problem or issue of concern (e.g. nutrients, metal loads, bioavailable metals)?
- what subsystem (including ecosystem type) should the model describe (e.g. freshwater, marine waters, estuarine waters, wetland, seagrass bed, mangroves)?
- which state should the model describe (e.g. base flow, flood event)?

Once formulated, the process model can be used to help define:

- the important components of the system and the important linkages;
- the key processes;
- the cause–effect relationships;
- the important questions to be addressed;
- the spatial boundaries;
- valid measurement parameters for the processes of concern; what to measure, and with what precision;
- site selection;
- the time and seasonal considerations.

2.4.1. Recognising the Key Processes

The monitoring team must aim to identify the key processes that define the 'cause and effect' of the system, and 'how the system works', because these are fundamental to the conceptual process model.

The major processes that affect water quality are broadly classified as hydrodynamic, physical, chemical and biological, and include:

- transport, flow, turbulence, flushing, mixing and stratification;
- precipitation, evaporation, wet and dry deposition;
- contaminant transport, sedimentation, burial, resuspension and diffusion;
- contaminant transformation, degradation, adsorption, desorption, precipitation, dissolution;
- sulfate reduction, methanogenesis, organic diagenesis;
- bioturbation, bioirrigation;
- organism growth, primary productivity, grazing, succession;
- nutrient recycling, loss, transformation, recycling, ammonification, nitrification, denitrification.

On the broadest scale, the monitoring team might be concerned with the sources and transport of contaminants, from a catchment to streams, rivers and estuaries. These form the basis of transport models, as shown for nutrients and metals in Figures 2.2 and 2.3.

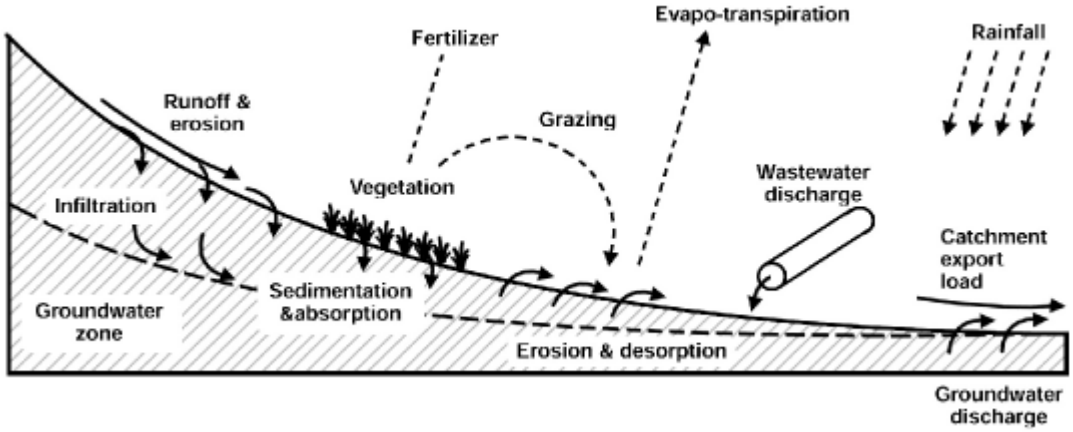


Figure 2.2. Model of sources and transport of nutrients through a landscape

The model in Figure 2.2 shows potential sources and transport of nutrients in the landscape. A more specific model might focus on a single water body, and the issue of concern in that water. Figure 2.4 illustrates a simplified model for phosphorus cycling in a stratified lake in relation to algal growth. The question that immediately arises is this: if you want to determine the extent of an algal bloom do you measure chlorophyll-*a*, algal cell counts in the water column, or some aspect of the scum? Further, if you wish to measure phosphorus concentrations in the water column, do you take the samples from the epilimnion or hypolimnion or both? (For this example, chlorophyll-*a* would probably be measured because it is a more reliable measure of algal biomass, and the samples would be taken from the epilimnion because this is where algal growth occurs.)

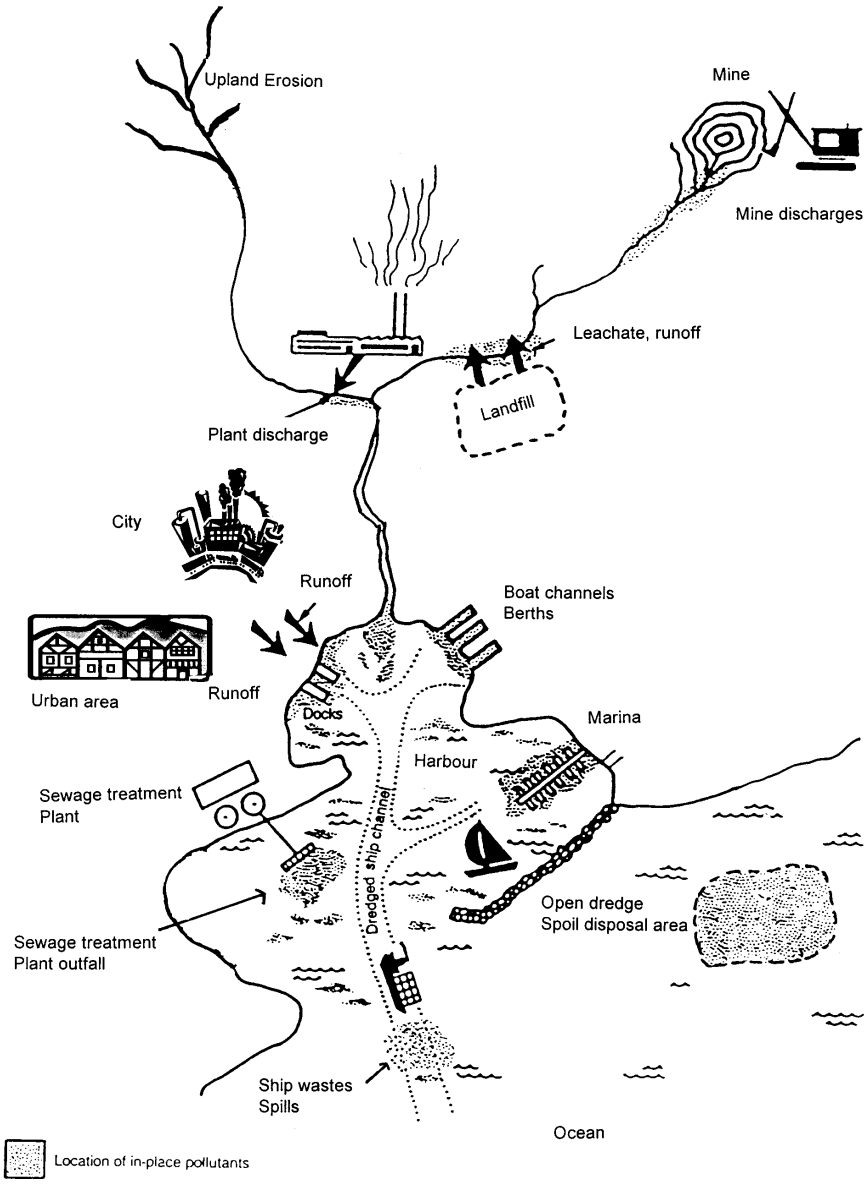


Figure 2.3. Model of sources of metal contaminants to the aquatic environment

Models can also include transformation processes — chemical, physical or biological. For example, Figure 2.5 illustrates the transformation processes that are associated with copper in a water body. Such models usually describe the chemical concentrations at thermodynamic equilibrium, and do not consider kinetics.

Kinetic models are based on the kinetics of reactions or growth, and are applicable when it is the rate of chemical reactions or biological growth that is important, rather than the thermodynamic equilibrium. These models can be used in describing the reactions of metals with particles, or biological growth processes such as the growth of algae or other organisms. The models are typically used for understanding oxygen utilisation, organism death and respiration, decay of pathogen populations, chemical and biological degradation of toxic substances, biodegradation of organic material, oxidation of organic and inorganic compounds, and excretion of toxic and non-toxic compounds by organisms.

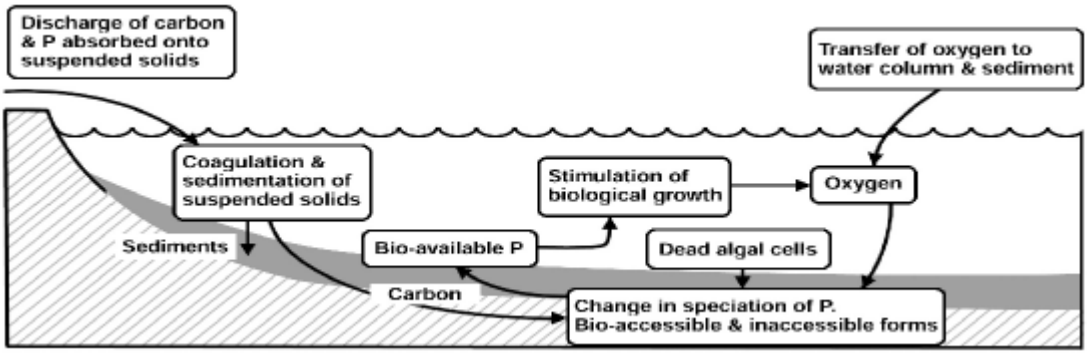


Figure 2.4. Model of in-lake or in-stream nutrient pathways

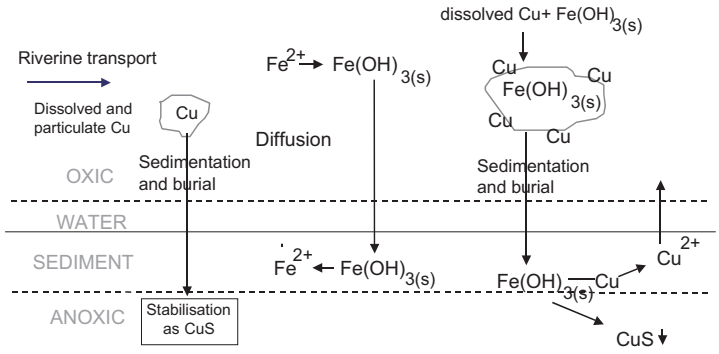


Figure 2.5. Model of pathways for copper in a water body

Models in the various climatic and geographic regions of Australia may differ. For example, conceptual process models will differ significantly between the wet tropics, the seasonally-dry monsoonal tropics, the arid interior, the temperate wet and the temperate Mediterranean regions. These factors can significantly affect study design, especially sampling strategies.

One of the limitations of using models is the assumption of continuity. In practice, the dominant processes may change as the previous state reaches some limiting condition. Different flow, mixing, chemical and redox regimes will turn alternate processes on and off.

It is important to be aware that the conceptual model being used might be wrong. Data that seem inconsistent can be important, leading to significant scientific breakthroughs from which new and more powerful conceptual models can evolve. The conceptual process models should be modified as information is collected and reviewed. The assumptions underlying the notional conceptual model should be validated and, if necessary, the model should be changed to reflect any new perspectives.

2.4.2. Testable Hypotheses and Conceptual Models

A monitoring objective is often framed as a testable hypothesis and based on a conceptual process model. This applies particularly to cause-and-effect studies, but a hypothesis can underpin monitoring for comparison with regulatory standards and even State of the Environment monitoring. Hypothesis testing is actually a test of the conceptual model.

Hypotheses usually take the form of statements or suppositions, such as these:

- variable A in a specified area or over a given time does not differ from a given baseline by more than some pre-defined difference;
- variable A in a specified area is not changed by more than some pre-defined amount per unit time;
- variable A (cause) is controlling variable B (effect).

Some hypotheses relevant to nutrient sampling might include these:

- phosphorus concentration is below (or above) the specified water quality guideline;
- phosphorus loading is controlling algal biomass;
- bioavailable phosphorus and nitrogen are limiting algal growth;
- phosphorus and nitrogen are being released from benthic sediments into the water column;
- in-flowing phosphorus is adsorbed by particles which settle to the bed of the lake;
- catchment activities have led to an increase in the annual phosphorus load to a lake.

A statistical hypothesis is a supposition based on available facts that can be subjected to a statistical evaluation, after further data have been obtained, to determine whether it can be accepted (or rejected). This sort of hypothesis is written in such a way that two outcomes are possible: either rejection or acceptance. The null hypothesis (that 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 when establishing hypotheses must be stated because their validity must be examined as part of the sampling design. If the hypothesis is rejected, the conceptual model should be refined.

There is some debate about the need to formulate a hypothesis. Monitoring is not always undertaken to overtly test some statistical hypothesis, although it almost always has a stated objective. Often, as Pratt (1976) noted,

> which hypothesis you are in is treated as overridingly more important than where you are in it. This is often an inappropriate view.

The requirement that monitoring be reduced to a null hypothesis and an alternative hypothesis is an artefact of classical statistical inference. Thus, hypothesis testing often forces the researcher to try to establish a significant difference between locations, say, instead of attempting to *describe* interesting spatial trends over a river reach.

The monitoring team must decide which of these approaches it will adopt in such cases because this will affect the data that need to be collected.

2.5. Setting Objectives

Once the monitoring team has defined the issue for monitoring, and has specified in general terms the information required from the monitoring program, and has agreed upon a conceptual process model, and, as a result, has further refined its understanding of the information that needs to be collected and why, it can finally write down a set of monitoring objectives.

Good monitoring objectives should be specific and precise, measurable, result-oriented, realistic and attainable, meaningful, concise and clear, and understandable. Clear objectives make it possible to design a sampling program to obtain the information required, but reviews of water quality

monitoring programs in Australia show that inadequate objectives are a common problem. The development of useful objectives requires practice and experience.

Typical objectives relating to nutrient dynamics and effects in aquatic systems might be these:

- to determine annual phosphorus loads to a specified lake from surface inflows, groundwater and sediment release (where the conceptual model has decided that all these sources are important);
- to determine the frequency of blue–green algal blooms in a number of specified water bodies over a defined period;
- to determine annual nutrient exports from a specified catchment to a specified river system.

A typical objective with respect to contaminants might be this:

- to determine if contaminant concentrations being released to a river under base flow from a specific industrial activity are exceeding the ANZECC & ARMCANZ water quality guideline trigger values for the protection of aquatic ecosystems in the receiving waters beyond the mixing zone.

Note that the objectives do not specify details such as sampling season or sampling frequency. Those are matters for the next stage, study design, described in Chapter 3.

Some examples of actual issues and resulting objectives are given in the four case studies in Appendix 4. For instance, an investigation of the aquifer that supplies groundwater to part of south-east Queensland was begun because the Logan–Albert catchment is being subjected to increasing pressure as a direct result of population growth. The objectives were to establish benchmark groundwater quality conditions for use in subsequent monitoring, to identify and understand the processes degrading groundwater quality in the aquifer, and to integrate the information obtained and provide advice to the responsible natural resource managers (see section A4.2.1). As another example, the major objective of a long-term monitoring program set up by the Great Barrier Reef Marine Park Authority in 1992 was to investigate the long-term trends and regional differences in nutrient status of the waters that comprise the world's largest reef ecosystem. In the last 140 years total nutrient input has increased by about 30% and this excess of nutrients has the long-term potential to damage the fragile ecosystem that exists within the Great Barrier Reef (see section A4.4.1).

The setting of objectives will commonly go beyond scientific issues by addressing management issues as well. This means that the resource manager needs to be involved in the negotiation of the monitoring program objectives. The resource manager must understand how the information to be collected will be used in the decision making process. If the only resources that the manager can make available are insufficient to meet the set objectives of the monitoring program, the program is not worth undertaking. The objectives may be rethought and more realistic objectives set.

Chapter Three
Study Design

3.1. Introduction

Once the monitoring team has accepted a conceptual model and defined the objectives of the monitoring program, the next stage involves general decisions about a more detailed design that also specifies data requirements. This is a fundamental stage that ensures that the sampling and analysis programs are cost-effective. It takes place before sample collection starts, and again involves interaction with the end-users of the information.

Figure 3.1 shows a framework for undertaking monitoring program design, and Table 3.1 is a checklist. The case studies in Appendix 4 illustrate the concepts of this chapter.

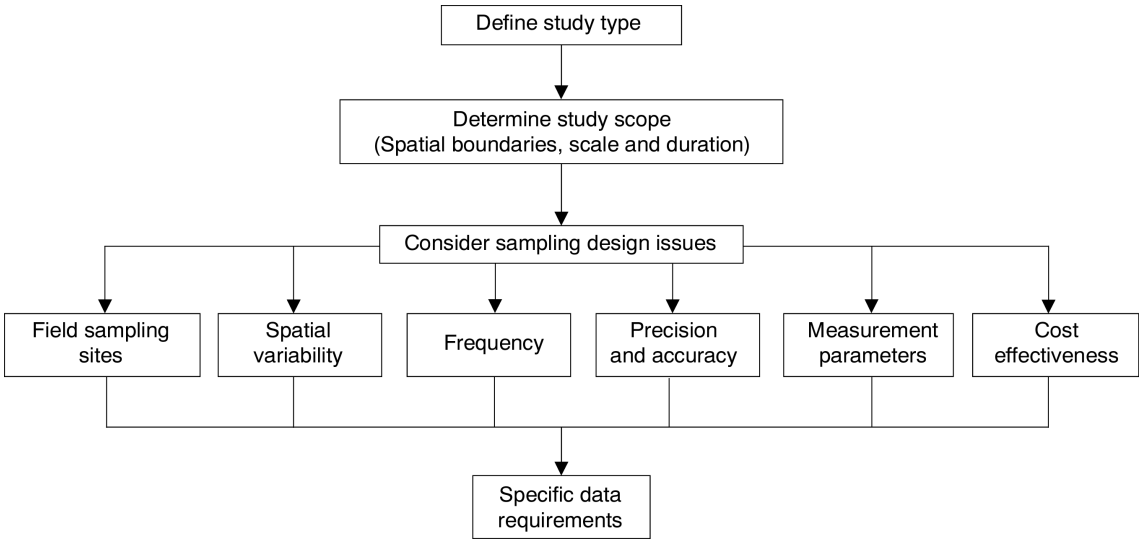


Figure 3.1. Framework for designing a monitoring study

3.2. Study Type

First the monitoring team must decide on the study type, because this will define the field sampling program and the path taken for subsequent data analyses. Three distinct study types can be identified:

- descriptive studies;
- studies that measure change;
- studies that improve system understanding (cause and effect).

Table 3.1. Checklist for designing a monitoring study

1.	Has the study type been made explicit and agreed upon?
2.	Have the spatial boundaries of the study been defined?
3.	Has the scale of the study been agreed to?
4.	Has the duration of the study been defined?
5.	Have the potential sources of variability been identified?
6.	Are there sufficient sampling stations to accommodate variability?
7.	Are the sites accessible and safe?
8.	Can sites be accurately identified?
9.	Has spatial variation in sites been considered, and have options to minimise this variation been considered?
10.	On what basis is the frequency of sampling proposed?
11.	Have decisions been made about the smallest differences or changes that need to be detected?
12.	Is replication adequate to obtain the desired level of precision in the data?
13.	Have the measurement parameters been chosen?
	(a) Are they relevant?
	(b) Do they have explanatory power?
	(c) Can they be used to detect changes and trends?
	(d) Can they be measured in a reliable, reproducible and cost-effective way?
	(e) Are the parameters appropriate for the time and spatial scales of the study?
14.	Has the cost-effectiveness of the study design been examined?
15.	Have the data requirements been summarised?

3.2.1. Descriptive Studies

Descriptive studies gather data to document the state of a system. They are the most basic of monitoring exercises. Typically they measure the spatial and sometimes temporal distributions of constituents within a water body for the purpose of (i) reconnaissance surveys, (ii) State of the Environment reporting, or (iii) assessing conformity with water quality guidelines or other agreed guidelines. They can determine background or baseline concentrations when a disturbance or development either is not expected or has not happened yet (baseline studies). In sediments, they can also identify changes that have occurred due to some earlier (historical) disturbance.

If the focus of the study has been descriptive, it is not usually possible to analyse the data subsequently to demonstrate causality. The need for this should therefore be determined in advance.

3.2.1.1. Baseline Studies

In baseline designs, no disturbance has occurred. An example of baseline designs is provided by some of the long-term water quality network programs that mainly monitor physical and chemical measurement parameters. Such programs are maintained so that they can detect or document any completely unanticipated changes in water quality. In these cases it is best to decide which measurement parameters to monitor, and the directions and sizes of changes or trends that would be important in those parameters (Green 1979; ANZECC & ARMCANZ 2000). When the monitoring team knows what changes to expect in the measurement parameter, it can refine the sampling design to avoid two very common pitfalls: either collecting insufficient data to detect the trend or change reliably, or collecting so much or such inappropriate data that ecologically trivial changes are detected. Well-designed baseline studies for which the likely nature of the disturbance can be anticipated are a prerequisite of the strong designs in section 3.2.2.

Baseline studies of sediments offer an opportunity to see the effects of historical changes in sediment contaminants, and can be used to establish the magnitude and perhaps timing of some past disturbances. The scientists and statisticians involved use their skills and insights to assemble information from appropriate studies. An example is the research program to detect increases in mercury levels in the Great Lakes (Green 1979). Because such studies are necessarily situation-specific, it is impossible to be prescriptive about the designs involved in them beyond noting that several independent lines of evidence strengthen any inferences about the effects of the disturbance.

3.2.2. Studies that Measure Change

When descriptive monitoring studies are repeated several times at the same locations, they can assess change. Such studies require relatively detailed planning so that locations can be identified and resampled. Data analyses can range from comparatively easy measurements of trends and simple correlations, to more complex evaluations that show if there has been a change of measurable significance. These are described in Chapter 6 and Appendix 5.

Monitoring is often done with the objective of evaluating the effects of a particular input or disturbance. If the timing and location of the disturbance are known, three categories of design are applicable (modified after Green 1979):

(i) *before–after, control–impact (BACI) designs*. Before the supposed effect occurs, two types of site can be identified: those that will be subjected to the disturbance and those that will not. The same parameter is monitored at both types of site before and after the disturbance to determine whether or not its *pattern of behaviour* over time at the disturbed site(s) changes relative to the control sites. After the disturbance starts, if the parameter's pattern of behaviour in the affected area(s) differs from its pattern of behaviour in the control areas, the differences are relatively unlikely to be due to chance.

(ii) *inference from change over time*. In this category of designs, no unaffected control site exists and change in a parameter can only be detected by comparison of data from one or more sites before and after the disturbance. In this design it is as likely as not that the change has occurred naturally over time, independently of the disturbance.

(iii) *inference from change over space.* In this category of designs, there are either unaffected control sites or there are sites that have been affected to varying degrees by the disturbance, but there are no valid comparable data collected before the disturbance. Sites used for the comparison may be upstream of the disturbed site, or on unaffected tributaries in river systems or estuaries, or in adjacent water bodies (e.g. wetlands, freshwater or saline lakes), or they may be distributed along some disturbance gradient (e.g. increasing distance from a point source). In such studies it is as likely as not that the values of the parameter in the disturbed and less disturbed sites differed before the disturbance.

Inferences should not be based solely on changes over time or changes over space unless there are no valid control sites or pre-disturbance data. Suitable spatial or temporal controls should always be used if they are available.

3.2.2.1. BACI Designs

The BACI designs have evolved in response to the common observation that the values of measurement parameters often differ naturally between any two ostensibly identical sites. The strongest versions of these designs base their inferences on interaction terms in a statistical analysis rather than on simple comparisons of means between sites. The logic of this procedure is best demonstrated first by discussion of Green's (1979) formulation of a BACI design, which is now regarded as the weakest of all the BACI designs, and then by an outline of subsequent improvements to the basic scheme.

Green (1979) proposed that environmental change would be detected if a measurement parameter were sampled from two separate sites, once before and once after a disturbance (Figure 3.2). One of

the sites would be the impact site (the site that would be subjected to the disturbance and potentially affected by it). The other site would be the control site, which would be similar in all relevant respects to the impact site except that it would not be subject to the disturbance. The sites would be chosen so that they were independent of each other with respect to the measurement parameter. If the impact site were affected by the disturbance, then, Green argued, this would be apparent in a significant interaction term in an analysis of variance (where the factors in the analysis would be 'time' with two levels, 'before' and 'after', and 'site' with two levels, 'control' and 'impact'). In graphical terms (Figure 3.2) the behaviour of the 'impact' site would change relative to the behaviour of the 'control' site after the disturbance. The values of the measurement parameter would not have to be identical in the two sites before the disturbance because the inference would be based on the interaction term in the analysis.

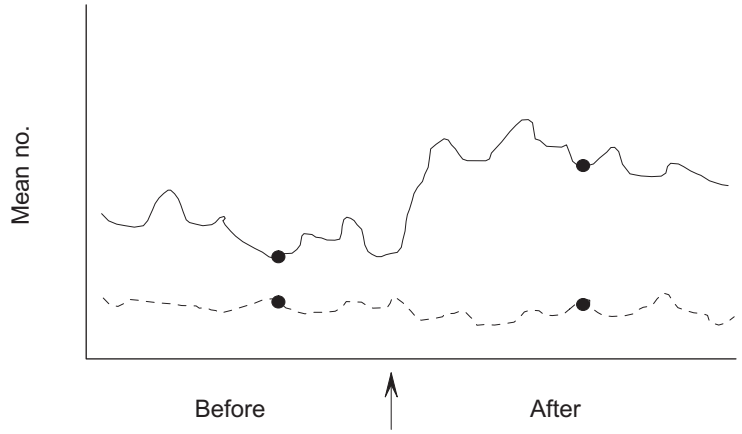


Figure 3.2. Illustration of a BACI design with single sampling events (denoted by dots) before and after a disturbance (arrow) in both a control site (dashed line) and an affected site (solid line) (modified from Underwood 1996)

Although Green's (1979) scheme was an important conceptual advance for environmental scientists, the notion of basing the inference of change on single sampling events from single sites of each type was criticised. The inference would be based exclusively on subsampling within each combination of site type and time (Hurlbert 1984; Stewart-Oaten et al. 1986); another site-specific disturbance event, unrelated to that being monitored, could confound the conclusions from such a design.

The preferred approach to circumvent this problem is to monitor more than one control site and to use multiple sampling events before and after the disturbance, as in the so-called MBACI designs (Keough and Mapstone 1995; Underwood 1996). The scheme is illustrated diagramatically in Figure 3.3. There are a number of important choices that need to be discussed when designing such programs, including the locations of sites, the number of 'before' sampling events, and the sampling effort required to model trends and dependencies through time. Although much of the literature about these designs focuses on analysis of variance, other statistical procedures (e.g. generalised linear models, see Appendix section A5.1.9) may be more appropriate and flexible for handling data that are not normally distributed. The data requirements of such procedures need to be discussed with a statistician before any data are collected.

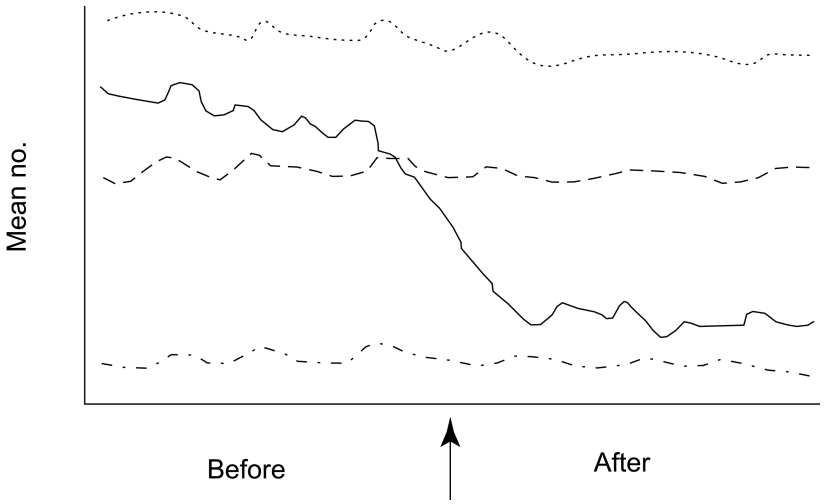


Figure 3.3. Illustration of Underwood's modified BACI design in which multiple random samples are taken before and after a disturbance (arrow) from three control sites (dashed lines) and an affected site (solid line) (modified from Underwood 1996)

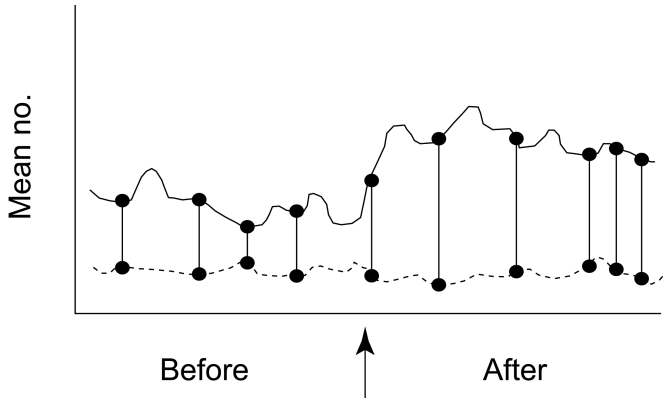


Figure 3.4. Illustration of a BACI design with multiple random samples taken before and after a disturbance (arrow) in both a control site (dashed line) and an affected site (solid line) (after Underwood 1996)

A number of variants of BACI designs have been proposed, and these are more fully discussed by their authors (Stewart-Oaten et al. 1986; Underwood 1991, 1992, 1994; Keough and Mapstone 1995, 1997). One commonly promoted variant applies to situations where there is a pair of sites: a single 'control' and a single 'impact' site sampled on many occasions before and after the disturbance (called BACIP by Stewart-Oaten et al. 1986) (Figure 3.4). For the inference to be strong from this design, the sites must be closely matched and some restrictive assumptions must be applied to the behaviour of the measurement parameter in the two sites. For example, if the measurement parameter were the abundance of fish, it would be presumptuous to assume that the population patterns and dynamics would necessarily be identical at the two sites. Thus, this approach should be used if only a single control site can be found, because localised site-specific events unrelated to the disturbance of interest can become confounded with the effect of most interest. Osenberg and Schmitt (1994) describe salutary examples of the problems inherent in these designs, for a marine system. The term randomised intervention analysis (RIA) has also been applied to this type of design (Carpenter et al. 1989).

3.2.2.2. Inference from Change Over Time

In some circumstances there are no suitable control areas, and so changes associated with a disturbance can only be inferred by comparing post-disturbance data with pre-disturbance data collected from the one site. Because there are no spatial controls, there is a chance that an unrelated disturbance may have coincided with the disturbance that is being monitored or assessed.

The main statistical procedures that can be used to analyse such data include (but are not restricted to) regression, trend, and time-series analyses. Sometimes the term 'intervention analysis' is used when time-series analysis is applied to a defined disturbance (e.g. Welsh and Stewart 1989). These procedures constitute a large and complicated area of applied statistics. Although some of the robust alternatives to the classical techniques are sketched in Appendix 5, it is beyond the scope of the Monitoring Guidelines to discuss them in detail. Expert statistical advice should be sought when planning and analysing these data, and particular attention should be paid to the modelling of interdependencies between successive sampling events and to choosing sampling intervals that are appropriate to the disturbance being monitored or assessed (e.g. Millard et al. 1985).

Often, these statistical procedures require data from a large number of sampling events and are most applicable to measurements of physical and chemical parameters (e.g. Welsh and Stewart 1989), although biological measurement parameters have been used in such designs (e.g. fish ventilation by Thompson et al. 1982). For such long-term designs, particular attention needs to be paid to coping with irregular sampling intervals and the inevitable missing data because classical statistical techniques are sensitive to both these occurrences (e.g. Galpin and Basson 1990).

3.2.2.3. Inference from Change Over Space

Often disturbances have already occurred or are alleged to have occurred, and scientists are required to judge the severity of impact or monitor the situation, either to assess whether recovery is occurring or to assess the success of remedial actions. Because such studies have no useful *pre-disturbance* data, inferences about the disturbance rely on spatial patterns. These patterns are found either in contrasts between disturbed and undisturbed sites or in sites chosen to represent a gradient of disturbance. The disadvantage of this class of design is that the observed pattern may be confounded with other environmental changes that are not related to the disturbance being monitored or assessed.

In rivers, to monitor recovery or dilution of the measurement parameter, it has been fairly common to select and sample a control site upstream of a disturbance and a series of sites downstream of the disturbance. Although this design is intuitively appealing, it has two problems. First, if sites are close together there may be inter-correlation between them that may mask changes (see also section 3.4.2). Second, there may be considerable natural variation in the measurement parameter that may not be captured in a single control site; therefore differences between the control and disturbed sites may not be due solely to the disturbance itself.

Multiple control sites, if they can be found, provide a stronger basis for inferring impacts resulting from a disturbance. A specialist statistician should be asked to clarify whether sites can be chosen to satisfy the assumptions of the analysis, or whether sufficient data are being collected to identify any spatial intercorrelations between the sites and allow valid inferences to be drawn.

Sometimes, it is not possible to find control sites that are undisturbed but resemble the disturbed site in all other important respects. Instead, *reference sites* are identified that are deemed to represent standards. Then values of the chosen measurement parameters at the disturbed site(s) are compared with values of the same parameters at the reference sites. This approach has been used for macroinvertebrate community structure in the AUSRIVAS procedure, and is explained in Box 1 (page 3-24). Alternatively, a gradient of disturbance can be identified either within the area surrounding the disturbed site (e.g. the seabed surrounding an oilrig) or across a number of sites across the landscape (e.g. a series of wetlands along a salinity gradient).

Gradients of disturbance — that is, values of measurement parameters that increase or decrease with distance from a point or boundary — are spatial patterns that are poorly described by classical

statistical techniques such as ANOVA (analysis of variance) and regression. Spatial statistical tools (e.g. Cressie 1993) should be more appropriate for describing these spatial patterns; for example, concentrations of toxicants in sediments, or abundance of species of benthic animals or plants. The classical spatial statistical tools can require very large areas to be sampled, and very large numbers of samples (Rossi et al. 1992).

Spatial analyses of data from sites that lie along a gradient of disturbance are sometimes termed *gradient analyses*. In these, some independent measure or surrogate of the disturbance (e.g. distance from source) is correlated with values of the biological measurement parameter. When an aspect of community structure is being measured, multivariate techniques such as ordination and clustering are used to relate the biological pattern to the spatial pattern (e.g. Warwick and Clarke 1993). These techniques are rapidly evolving, and novel ways are being developed to quantify the relationships between multivariate biological responses (as expressed by dissimilarity measures) and spatial patterns.

3.2.3. Studies for System Understanding

Some studies are made with the aim of finding out more about a particular system; for example, to better understand aquatic ecosystems and the physical, chemical and biological processes that operate in them. A deeper understanding may reveal relationships among the variables operating in the system, enabling predictions to be made about the behaviour of the system in situations beyond existing data and experience.

If the objective of a study is to establish cause and effect relationships, the sampling program must be designed for this purpose from the start. For this objective, the monitoring team may need to run additional experimental studies in which they can manipulate the system in a controlled manner and measure the system's response. In this case the sampling regime must be designed so that at least one of the potential outcomes is unequivocal. Manipulative experiments are routinely conducted in laboratories, but in the field they can be expensive and it may be impossible to control all the confounding variables adequately.

In studies of cause and effect, even the best experimental or survey design may be insufficient by itself. No design can completely defend against all unidentified confounding influences (Stewart-Oaten et al. 1986; Eberhardt and Thomas 1991; Underwood 1994). To establish cause and effect, therefore, the monitoring team must assemble independent lines of evidence and circumvent the potential for inferential problems akin to those faced by epidemiologists. Beyers (1998) has attempted to combine epidemiological criteria (Hill 1965) with postulates for environmental toxicology (Suter 1993) (Table 3.2). Not all of these criteria need to be met, but strength, consistency and specificity provide the strongest evidence for causation. Where a disturbance is chemical, indicators of exposure (e.g. contaminant concentrations in tissues) also provide strong evidence for causation. Whether Beyers's emphases are appropriate or not is likely to be debated as investigators try to formalise the ways in which they combine evidence in environmental studies.

It should be noted that the results from a study that measures change also contribute to system understanding by demonstrating a link between a particular human activity and a specified effect in the system under consideration. However, they do not establish cause and effect. Some other unknown cause may have resulted in the effect. To establish a cause–effect relationship some characteristic of the activity needs to be linked to the change observed.

However, studies for improving system understanding are not always done to show cause and effect. For example, the conceptual process models in Figure 2.4 and Figure 2.5 outline a system understanding with respect to nutrients and copper, respectively. A study could monitor the changing significance of processes in these models, over space and time.

3.3. Scope of a Study

When it defined the objective of the monitoring study (see Chapter 2), the monitoring team would have identified the study site in broad terms, e.g. the River Murray, or Sydney Harbour, or the Brisbane River and Moreton Bay. Now the monitoring team can set the spatial boundaries of the study and consider questions of scale and duration.

3.3.1. Spatial Boundaries

The setting of spatial boundaries is important because inappropriate boundaries might focus the study away from important driving or consequential factors. This decision must be based on the issue of concern and the ecosystem rather than on convenience and budgets. In an investigation of effects of catchment activities on rivers, lakes and estuaries, for example, the spatial boundaries would normally be those of the catchment.

As an example, for a study of the Brisbane River and Moreton Bay, the monitoring team might consider if the study should look only at these water bodies, or if it should include tributary creeks or the broader catchment, or if the study should be restricted to the major receiving waters.

The pertinent point here is that the monitoring team needs to explain the logic behind its decisions with respect to the spatial boundaries of the study.

3.3.2. Scale

Scale refers to the spatial and temporal ranges over which a system is observed, i.e. the appropriate level of resolution to answer the questions of concern. Different processes operate at different scales. For example, the movement of sediment in a river system may take tens of years at the catchment scale, toxicant effects may occur over days and may be localised, while nutrient enrichment may occur over kilometres and the response may take weeks.

The scale of the study should be chosen in relation to the study's objectives after the monitoring team has considered the measurement opportunities at the various possible scales and the likelihood of collecting reliable and valid measurements. The cost of data collection at the various scales should also be assessed.

Will the necessary measurement parameters be spatially uniform? As the spatial extent of data collection gets larger, so the distribution of the measurement parameters can become more heterogeneous and patchy, and more replicate samples can be needed to achieve the same confidence in the results. It is essential to choose an appropriate scale relative to the phenomenon under consideration and then sample at that scale.

3.3.3. Study Duration

Similar problems affect the decision about how long a study should last to address the issue of concern. Given the variability of natural rainfall and hence streamflow, what length of time might be required to achieve an appropriate understanding of the system?

The appropriate length of the study is an important decision. Few hydrologists would make definitive statements on the quantity of water resources with data from only two or three years, yet frequently conclusions from water quality studies are expected from less than three years' data. What is a reasonable duration for the study? How long will it take for a sufficient variety of rainfall events (from droughts to floods) to be experienced to allow the monitoring team to study the system under extremes?

Table 3.2. Criteria to formalise the use of independent lines of evidence in inferring causation in impact studies[a]

Name of criterion	Description of criterion	Example[b]
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 measurement parameter 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, a decrease in density of the organism is not diagnostic of the disturbance because the population density of this organism may be reduced by other, natural, processes
Presence of stressor in tissues	Measurement parameters of exposure (e.g. residues, breakdown products) must be present in the 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 spillages of the toxicant are usually followed by sharp declines in the density of the organism
Biological gradient	A dose–response relationship exists (i.e. response of measurement parameter is a function of increases in magnitude of disturbance)	Laboratory toxicology 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 chemicals known 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 experiments with different but related species

[a] From Beyers (1998)

[b] A hypothetical example of the response of biological measurement parameters to a toxicant, as an illustration

3.4. Sampling Design

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

Before a field sampling program can be planned, some idea of the expected spatial and temporal variability of the measurement parameters is needed. Often, general information about variability can be obtained from published work during the formulation of the conceptual process model of the system. For example, oxygen concentrations are known to vary diurnally and to differ between the epiliminon and hypoliminon in lakes; phosphorus, bound to sediment, is known to be transported during rainfall events. Typical types of variation are caused by:

- spatial variability because the environment is heterogeneous;
- time dependence, temporal, seasonal effects;
- disruptive processes;
- dispersion of chemical contaminants.

Normally, the design of any investigation, and particularly a monitoring program that is to be ongoing, cannot be settled without a pilot study. This short period of intensive monitoring outlines the nature of the prevailing system and particularly its temporal and spatial variability. Then the monitoring team can choose a sampling regime and frequency that should provide a representative profile of the system for each measurement parameter and piece of information required. It can decide on appropriate numbers of replicate samples to provide the precision required for the statistical analyses used in the study.

The pattern of sampling in space and time is of critical importance. Although most statistical techniques require random sampling, simple random sampling is sometimes difficult to achieve, and may not be cost- efficient. The main patterns of sampling are outlined briefly here, but they are more fully discussed in basic texts such as Cochran (1977). A major problem during sampling is representativeness. The intellectual challenge is to design a sampling approach that minimises errors. The errors in accurately representing a water body or population by a sample, and a sample by a sub-sample, can far exceed errors in analysis (Gy 1986). These and other statistical sampling issues are reviewed by Helsel and Hirsch (1992).

3.4.1. Patterns of Sampling

3.4.1.1. Simple Random Sampling

The basic requirement of most statistical procedures is that each sample unit in the population of interest has an equal probability of being selected and included in the sample. There should be no conscious or unconscious selection of units to be included in the sample. A computer-generated set of random numbers can be used, but this usually requires that a grid or co-ordinate system be established in the study site so that each potential sample member can be identified. This, in itself, can be logistically difficult. So called 'haphazard sampling', in which sample units are selected without the help of random number tables, is *not* random. Procedures such as throwing a quadrat over one's shoulder or sticking a pin in a site map while blindfolded are subject to unconscious bias, which will result in biased estimates. Random sampling is discussed in more detail by Thompson (1992).

Simple random sampling may not be the most cost-efficient sampling pattern because of variation within the site or time period of interest.

3.4.1.2. Stratified Random Sampling

Stratified random sampling can often be substantially more efficient than random sampling; it is typically used in audit monitoring or to compare water quality against a guideline value. In stratified random sampling, the system to be sampled is divided into parts (strata) in each of which the variable of interest is as uniform as possible. Strata need not be of equal size. The numbers of sample units allocated to each stratum can be either in proportion to the size (area, volume) of each stratum or in proportion to the variance within each stratum.

Strata may be spatial or temporal. For example, for water sampling to measure nutrients, chlorophyll and algae, a lake can be divided spatially into the epilimnion and hypolimnion, or an estuary can be stratified on the basis of a salinity gradient. Temporally, if nutrients are more variable in one season than another, more sampling effort can be allocated to the most variable season, particularly if estimates of the annual concentration or load of the nutrients are the focus of the program. Sometimes strata result from an interaction of spatial and temporal processes. For example, suppose fish in a lake are being collected to study the accumulation of chemical contaminants, it is important to consider fish mobility and fish age (size). Older fish often accumulate more of a contaminant. Fish ages (sizes) then become the sampling strata, instead of geographical locations or particular periods of time.

3.4.1.3. Systematic Sampling

In systematic sampling, sample units are collected at regular intervals in space or time. When properly planned and executed, systematic sampling can be as unbiased as random sampling, and can be significantly cheaper (see Cochran 1977 for a full discussion). However, care needs to be exercised to ensure that bias is not inadvertently incorporated into the sampling scheme. For instance, regular sampling schedules may coincide with periodicities in the disturbance being monitored (e.g. discharges from a factory may be consistently lower in the morning and greatest just before shut-down in the late afternoon). Similar situations can arise spatially.

There needs to be a good descriptive base of background information so that systematic sampling can be both cost-effective and unbiased, and it is essential to document the assumptions and choices made when executing such a sampling regime.

3.4.2. Selection of Sampling Sites

It is important to select sites that provide appropriate spatial information. The problem being addressed will largely determine the general locations of sampling sites. The statistical analyses that will be used to interpret results (see Chapter 6) will also guide this decision (Ward et al. 1990). When ecological impacts are being assessed, sites will normally be located relative to the likely disturbance. Only rarely will sites be located randomly, as discussed above (section 3.4.1), but when this is done the number of sites and the extent of homogeneous areas in which they may be located can be determined from the pilot study. Multivariate classification procedures can be used for grouping similar sites, to define homogenous areas (Clarke and Warwick 1994).

When selecting appropriate sampling locations, the monitoring team should consider the possibility of seasonal variations and of local variations in other parameters to be measured (e.g. sources of contaminants), by referring not only to the pilot study but also to past records. These could be records of activities in the catchment, aerial photographs, plans and maps of land use, and oral or other records of the sites and the catchments under investigation. The team may find, for instance, that the water quality should be monitored not only in major surface waters and groundwaters that might receive inputs of substances from diffuse sources, but also in small creeks hydrologically connected with those waters.

It is also important that sites be selected to minimise any artefacts from human interventions that are not part of the monitoring program. For example, flow may be modified around jetties or bridges and that may affect some benthic measurement parameters, resulting in spurious data if the effect of the jetty or bridge is not the focus of the monitoring program. Similarly, weirs and similar structures in

rivers often alter both the flow and the chemical conditions, and sampling sites need to be located far enough up- or downstream of such structures if the water quality of the free-flowing water is the major focus of the monitoring program. In the field, the actual sampling sites will usually be selected by personal judgment using pragmatic considerations such as accessibility and safety.

When control or reference sites are included in the design, care needs to be taken to ensure that they are closely matched with the site being assessed. Sometimes information on covariates can be collected at all the sites and used to adjust the values of measurement parameters for inherent differences between the sites; the assumptions of the statistical analyses of such data need to be met (see Chapter 6). For example, in studies of metals in sediments, sediments from the reference site should have grain size and organic content similar to those from the test sites.

If sampling sites are too close together, or samples are collected at too close a time interval, autocorrelation or serial correlation between sites can invalidate the assumptions of independence made in some classical statistical designs. What constitutes *too close* (spatially or temporally) depends both on the nature of the measurement parameter and the dispersion of the contaminant. 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.

Some water quality programs (usually based on chemical and physical measurement parameters) rely on networks of sites. There are high costs associated with monitoring, so monitoring programs should be optimised with regard to networks and sampling. A number of 'spatially optimum water sampling plans' exist, and their merits have been reviewed (Dixon and Chiswell 1996).

Finally, some pragmatic considerations for selecting sites.

- Safe access must be ensured under all conditions. If the sites are inaccessible during the wet season, for example, then the monitoring program cannot address questions about water quality during wet seasons.
- Sites also need to be accurately identifiable so that they can be sampled repeatedly. Global positioning systems greatly ease this task in areas of low relief and in off-shore marine environments.
- Groundwater quality monitoring programs often require a carefully staged approach, taking account of local geology, the vulnerability of the aquifers to contaminants and land use patterns, and any changes in pattern.

3.4.2.1. Spatial Variation Within a Sampling Site

There may be spatial variation *within* a site that needs to be quantified in the monitoring program, because otherwise the estimates of the chosen measurement parameter may be imprecise or even inaccurate. For example, in thermally stratified waters the depth of sampling is important because the concentrations of many measurement parameters (e.g. hydrogen ions, dissolved oxygen, nitrate, hydrogen sulfide, plankton) can vary greatly between the top and bottom layers. In rivers, samples taken from the edge rather than from mid-stream are likely to contain quite different amounts of suspended material and therefore different amounts of various compounds bound to the particulate matter. In benthic sampling for biological parameters (e.g. invertebrates, algae) or for sediments, the habitats or sediment types may vary at a site. In formal terms, these different habitats or water types within a site are called *strata*.

It is important that the monitoring team recognises that stratification in the measurement parameter will affect the data being obtained. There are three options for dealing with such strata:

- restrict the scope of the inference to a particular stratum. For example, if sandy sediments dominate the substrate at all the study sites, it may be sensible to confine sampling to sandy substrates. The stakeholders must be made aware that the inferences drawn are applicable only to 'sandy substrates within the sites' and cannot be generalised to strata that were not sampled within the sites.

- divide the sampling effort among the strata. Here the goal is to estimate the value of the measurement parameter for each site as a whole rather than for a stratum within the site. Stratified random sampling (see section 3.4.1.2) is an example of this procedure that is fully explained in basic texts (e.g. Cochran 1977; Elliott 1977). The number of sample units allocated to each stratum can be determined by the relative sizes (e.g. area or volume) of each stratum, or by the within-stratum variation of the measurement parameter(s).
- make separate estimates for each stratum (if this is consistent with the study objectives). Here the monitoring team may want to identify the nutrients in each stratum. For example, at each site in a reservoir separate nutrient samples can be taken from the epilimnion and the hypolimnion (i.e. two strata). They are then kept separate throughout the analyses.

When measurement parameters are being sampled in the water column, it is sometimes assumed that the water is well mixed and that a mid-water or mid-stream sample will be sufficiently representative. This may not be the case. Even in fast-flowing mountain streams, water can be observed flowing upstream in eddies. In larger rivers, tributary water may not mix fully with the mainstream for many hundreds of metres or even kilometres. In estuarine waters, salinity may be significantly stratified, and all water bodies can have gradients of redox potential and temperature. Even if the monitoring goal is just to measure the average concentration of a chemical in the water at a site, the sampling process must be planned so that the within-site variation is included in the estimate.

The same situation apples to the monitoring of aquifers, where groundwater quality is almost always stratified vertically, and where there can also be significant lateral variation in quality (e.g. in areas where there are multiple point sources or variable diffuse sources of contamination). There is much less dispersion of contaminants in groundwater than in surface waters, and so natural spatial variability is potentially much greater than in surface waters.

3.4.3. Sampling Frequency

The objectives of the monitoring program dictate the basis for determining sampling frequency. Thus a program to detect conformity with a guideline might be based on daily, weekly or quarterly sampling. The monitoring team must decide whether seasons are important, e.g. wet vs. dry season in the tropics, winter vs. summer in temperate regions where snow melts might be important.

Patterns in time include natural systematic changes, ranging from tidal cycles to larger scale events such as the El Niño–Southern Oscillation. These may be periodic and predictable (e.g. tides, seasonal filling of a wetland) or non-periodic (e.g. storms or floods in streams). Within these events, there are unpredictable variations (e.g. changes in the recruitment of a species after minor natural disturbances; changes in the concentrations of chemical measurement parameters after rainstorms).

Non-periodic events such as storms and associated runoff can have a dramatic impact on water quality that might be missed by sampling on a fixed time interval. If the monitoring team identifies this possibility during its preliminary assessments, it must design a sampling program that includes these events. Rapid changes in flow can profoundly affect the concentrations of measurement parameters and therefore the representativeness of sampling. Even under relatively stable flow conditions, the team must measure hydrological parameters when water sampling if the program is measuring the loads of a measurement parameter rather than its concentration.

The monitoring team must understand the system and the problem and issue being investigated, as illustrated by its conceptual process model, before it can select appropriate time intervals for sampling. The program objective and the expected statistical analyses can both influence the time interval chosen between samples. For example, the team may want to be 95% certain of detecting a 5% increase in nutrient levels. Once this objective is set, it will be a relatively straightforward statistical matter to determine the frequency of sampling.

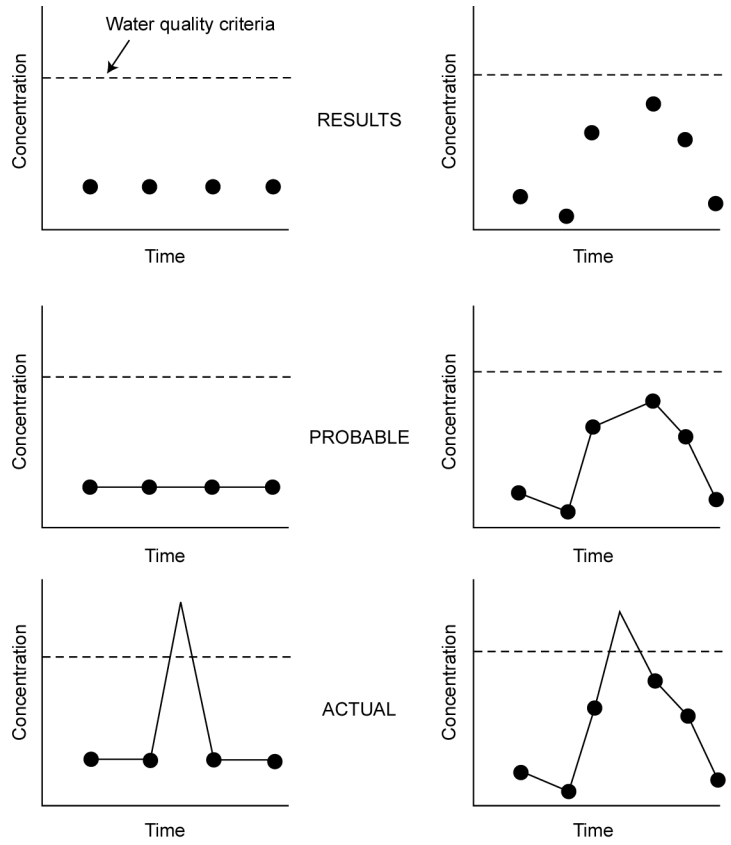


Figure 3.5. Frequency of sampling: interpretations of sampling data (modified from Maher et al. 1994)

The values of a particular measurement parameter may not vary at all time scales. If a measurement parameter has a predictable temporal pattern (e.g. recruitment with onset of the wet season, or deoxygenation during thermal stratification), the monitoring program must sample this measurement parameter at a frequency that suits this periodicity. Then trends can be estimated, especially in chronic or 'press' impacts. If a disturbance is only likely to take place at a certain time of year, for example mine wastewater discharge during the wet season in Kakadu (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 (e.g. stormwater discharges), the monitoring program must sample at several time scales. Some measurement parameters give snapshots of immediate condition; some are integrating measures that reflect conditions over the past (x) months. These time-scale decisions need to be based on:

- the characteristics of the parameter being measured;
- the purpose of the data collection;
- the statistical or other tools that will be used to interpret the data; for instance, for time series analysis the monitoring team may have to decide on and set a definite sampling interval;
- the characteristics of the response of interest; for example, weekly measurements might be appropriate for measuring the development of an algal bloom but not for investigating fish. The generation time of the organism might be the critical determinant of time scales.
- recognition that a process cannot be measured if it takes longer to happen than the period over which measurements are made.

The case study of the Great Barrier Reef (see Appendix section A4.4.2) illustrates decisions about sampling frequency.

The frequency of sampling is especially important when the objective of the monitoring program is to ensure that particular guidelines or standards are not exceeded. Figure 3.5 shows some possible mis-interpretations that arise from sampling at inadequate frequencies. The data values obtained at the selected sampling frequency are all below the water quality guideline. Actually, there were excursions above the guideline between samplings that were not evident at the selected sampling frequency. Mathematical formulae are available to calculate the sampling frequency required for a particular study (Sharp 1971; Montgomery and Hart 1974), but these are not in widespread use.

3.4.3.1. Specific Concerns for Biological Measurement Parameters

Biological sampling should also take into account the time-dependence of an organism's behaviour. For example, Magmann (1991) re-examined a published study on the Northern Red Belly (*Phoxinus eos* and *Phoxinus neogaeus*) in which the densities of both fish species were reported to be at their highest at or near the shore. The original conclusion, based on a 16–18 hour trapping period beginning at 1600–1900 hours, was that both species exploited the same microhabitat. However, the initial study failed to recognise that *Phoxinus eos* has a diurnal pattern of inshore–offshore migratory behaviour. These 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 sampling interval (3–4 hours) was required to observe this movement. The density of fish offshore seemed to be lower because the fish shoals had broken up. Unfortunately there are few behavioural data for Australian species, so guidance is difficult to obtain.

Biological parameters may have the problem of serial correlation because of the long life-span of the organisms involved. The size of fish populations, for example, may depend on year-to-year variations in recruitment that may not be consistent across all the sites included in a study. Auxiliary data on the age-structure of the populations would be necessary to unravel these effects. Serial correlation will cause problems if the statistical methods being used assume that the measurements at different times are independent. Chapter 6 gives further discussion of these concerns.

3.4.3.2. Specific Concerns for Chemical and Physical Measurement Parameters

Special care should be taken when measuring some chemical and physical parameters, e.g. dissolved oxygen and pH levels in still or slow-moving surface waters. The values can change dramatically in these waters during the day, through photosynthesis and respiration. For example, dissolved oxygen must be measured before sunrise to obtain the diurnal minimum; diurnal pH fluctuations occur when carbon dioxide concentrations vary, and the pH decreases at night when dissolved carbon dioxide and carbonate accumulate in the absence of photosynthesis. The practice of sampling at a certain time of day, without regard to the changes that occur between daylight and darkness, can therefore result in misleading data.

If concentration measurements are being used to calculate loads, it will be important to decide how to relate flow and concentrations, and on what time basis. There are four common types of system with differing dominant processes that must be recognised when considering the frequency of sampling:

- base flow and point source discharges are the major determinants of water quality; or
- runoff (volume of storm event) and non-point sources are the major determinants of water quality; or
- remobilisation is the major determinant of water quality; or
- diurnal cycles (tidal cycles or biological activity) are important.

Australian freshwaters are characterised by highly variable flow, so flow is a major issue for them. It affects both water quality and biology. Changes in flow can alter water quality parameters rapidly and sometimes unpredictably for several reasons including these:

- hydrological changes alter the relative proportions of discharge originating from runoff, baseflow and groundwater. Runoff water may be of better quality than groundwater and baseflow (Hart et al. 1987) but is not invariably so. This can often be observed as a period of decreasing electrical

conductivity during the rising limb of the hydrograph, followed by rising electrical conductivity during the falling limb of the hydrograph. Alternatively, runoff may contain increased concentrations of nutrients (from fertilised fields, urban areas or sewage treatment plants) or heavy metals and organic compounds (from contaminated sites).

- rainfall events within a catchment may give different patterns of water quality depending on their locations;
- deliberate releases of contaminants may coincide with extreme hydrological conditions to take advantage of the large dilution factors available then;
- extreme rainfall conditions may breach bunds and other containment devices used for the retention of contaminants;
- in the case of temporary water courses, very large changes in water quality may occur during extreme recessional flow and during the 'first flush' of new flow. In the latter case, chemical species that have accumulated in catchment soils and near-surface groundwater (sometimes acidic from organic degradation or sulfide oxidation) may dramatically alter the concentrations of some measurement parameters.

It is well known that in many water bodies the total heavy metal and phosphorus concentrations are correlated with flow discharge, particularly during the early part of a stormwater runoff or flood event. At this time there are more suspended solids and correspondingly higher concentrations of associated heavy metals and phosphorus. Therefore if a river is only sampled at base flow for chemical and physical measurement parameters, the resulting data will not truly represent the natural range of heavy metal and phosphorus concentrations in the water body. Similarly, in the initial stages of an algal bloom, the numbers of algal cells may double every 2–3 days. If the monitoring program is measuring some aspect of nutrient fluxes, the sampling needs to reflect flow events that transport materials into and through the aquatic system. Catchment exports to rivers and estuaries are also assessed by intensive sampling during events.

Traditionally, the designers of monitoring programs have not considered sampling under different flow regimes. However, in the majority of Australian rivers, most (70–90%) of the annual flow and constituents are discharged under high flow or event conditions even though these may prevail for only 1–10% of the time. Under these conditions, the dominant water quality processes are the transport and deposition of discharged material during the flow event, followed by in-stream remobilisation of deposited material in the 10–30 days following the event.

Where the issues underlying the monitoring program relate to flow, the monitoring team must consider the following:

- 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;
- 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.

To solve the difficulty of sampling at all flow regimes, a range of robust and reliable automatic sampling devices can now be obtained. These are capable of being automatically triggered by rising flow, and can automatically take samples at predetermined times or stream heights, to provide a comprehensive picture of changing constituent levels throughout an event. These data can then be used in association with flow data to calculate the contaminant export (or load) for a storm or flood event. Monitoring effort can be reduced over the long periods between events because there is little water quality variation during these conditions. An exception to this condition is the case of base flow with point source discharges.

3.4.4. Sample Numbers and Precision

An important aspect of the sample design is the number of samples to be collected to address the monitoring program's objective. This will largely depend on the nature of the investigation. In descriptive studies or in studies to determine cause and effect, the number of samples will determine the power of the data to assess differences. In studies that measure change, there must be enough samples to detect the minimum effect, or smallest differences or changes, that will cause management action — the 'effect size' (Keough and Mapstone 1995 p.102).

The monitoring team must decide on the required precision and accuracy. How many samples are needed for measuring each parameter at each site precisely on each sampling occasion; how many samples can the monitoring program afford to take? The team will base its decision on the results of the pilot study or on other reliable estimates of the variance and the costs of sampling (Keough and Mapstone 1995). The appropriate level of replication is not a simple decision (Segar et al. 1987; Mapstone 1995) because it must:

- be scientifically attainable;
- be attainable through a sampling and analysis program which can be accomplished in a cost-effective manner;
- minimise the risks of falsely detecting a disturbance or environmental impact when one has not occurred (giving a false alarm), or alternatively missing an environmental impact if it has occurred (giving a false sense of security);
- detect differences or changes that are environmentally important — that is, the change must have ecological meaning to the system of concern.

The smallest differences or changes that must be detected determine the number of spatial and temporal replicates needed (Norris and Georges 1986) and the precision needed. If a copper guideline concentration is 5 µg/L, is it important, environmentally, to be able to detect 5.01, 5.1 or 5.5 µg/L? This not the same as statistical significance (see Appendix 5 for a fuller explanation; see Mapstone 1995 for an explanation in a hypothesis-testing framework).

Once the difficult scientific and socioeconomic questions have been answered about the size of the differences or magnitude of the trends that must be detected, the number of replicates required can be calculated (see Appendix section A5.1.10). This is effectively an application of statistical power analysis, and more detailed explanations of the basics are found in many introductory texts (e.g. Cohen 1988; Sokal and Rohlf 1995). Various formulae are available for calculating the required numbers of replicates (e.g. Norris et al. 1992; Keough and Mapstone 1995), although investigators should be aware of the distributional assumptions behind such formulae (see section 6.3.5). On the other hand, decisions about optimum sample sizes for complex designs may not be easy (Green 1989, 1994; Norris et al. 1992; Keough and Mapstone 1995, 1997), and professional statistical assistance may be required.

Two related issues need to be borne in mind during this process of determining the number of replicates required for the program. First, investigators and their statistical consultants need to be clear about what constitutes a true replicate for the question being addressed by the monitoring program; 'pseudoreplication', where there is autocorrelation between apparent replicates, has been rife in many environmental programs (Hurlbert 1984; Eberhardt and Thomas 1991). Second, many programs will require sub-sampling within sites and within time periods to improve the precision of estimates.

There are trade-offs with costs, but unless the sampling is done in a way that enables the required data to be collected it cannot hope to answer the study objectives. If the monitoring team finds that resources are limiting, they may need to reconsider the sampling objectives. The monitoring team will need to decide on the sampling effort that is required to test critical hypotheses, if these are being used. If precision will be below that at which the critical hypotheses can be tested, the proposed sampling design is a waste of time and money. If the information generated by the monitoring program 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 as political or social costs.

3.5. Selection of Measurement Parameters

The selection of measurement parameters is a vital element of the monitoring program design. A wide range of physical, chemical, ecotoxicological and ecological measurement parameters can be used to provide information on water quality. There is no simple or single physical or chemical measurement parameter that defines the quality of water. The choice of measurement parameters depends on the values ('environmental values') assigned to the water body (ecosystems, drinking water, recreation, industry, agriculture, aquaculture), and therefore on the objectives of the study. Furthermore, the guideline values acceptable for a measurement parameter for a particular use can differ geographically and temporally (ANZECC & ARMCANZ 2000).

The Water Quality Guidelines promotes the idea of integrated assessment. This approach merges biological (effects) and chemical (causes) approaches, and combines holistic field evaluations of impacts at the community and population level with laboratory toxicity tests. While the order of importance of biological versus chemical and physical monitoring can be debated, all provide important information as part of the integrated assessment of ecosystem health. Three-pronged studies (the triad approach), using chemistry, ecotoxicology and ecology have been promoted for sediments, and apply equally to waters (Chapman 1990).

Chemical measurements provide concentrations of specific contaminants that might be the cause of specific effects or modifiers of them. Biological assessment can be broadly subdivided into laboratory studies of chronic and acute impacts on individual species (ecotoxicology), and field measurements of structure, populations of species and their diversity, and function (aquatic ecology). Ecotoxicological and ecological measurements are non-specific, responding to the sum of the contaminants in the system. They integrate the effects of these contaminants over time and provide a more direct measure of the health of an aquatic ecosystem. Some taxa appear to be extremely susceptible to certain chemical contaminants and so provide a sensitive tool and early warning system for detecting slight contamination.

Table 3.3. Checklist for selection of measurement parameters[a]

Relevance	Does the measurement parameter reflect directly on the issue of concern?
Validity	Does the measurement parameter respond to changes in the environment and have some explanatory power?
Diagnostic value	The measurement parameter 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?

[a] Adapted from Maher and Cullen (1997)

The monitoring team's conceptual process model has already defined the water body and the problem and underlying issues that it is monitoring. Now the team must decide whether to measure the driving or causal factors (e.g. contaminant concentrations), or the consequential or resultant factors (e.g. effects such as toxicity, or algal biomass), or both — and why. The fundamental questions are these: how will the two sets of data be used; will the chosen measurement parameters have relevance to the problem or issue; will they change (or react to change) within the time-frame of the monitoring program; are they readily measurable? Table 3.3 is a checklist of these and other characteristics for selecting appropriate measurement parameters. These considerations have led to an upsurge in the

monitoring of biological outcomes rather than chemical inputs to a system, sometimes in community-based programs such as the algal alert programs and the biological components of Waterwatch.

A trade-off may be required between the exactitude of some measure and its cost or difficulty of measurement. The monitoring team might wish to know the concentrations of dissolved or bioavailable contaminants (e.g. phosphorus, metals) but may settle for total concentrations because they are easier to measure and more reliable (Lambert et al. 1992).

In many studies, parameters are measured that are not related to the conceptual process model of the system on which the study is based and for which no predictive power has been assumed. The reasons for including these measurements need to be justified.

The Monitoring Guidelines does not recommend any particular biological or physical or chemical measurements. For detailed description of physical, chemical and biological monitoring approaches, including ecotoxicology, see the Water Quality Guidelines (ANZECC & ARMCANZ 2000).

3.5.1. Physical and Chemical Measurement Parameters

The physical and chemical measurement parameters used for assessing water quality are discussed in detail in the Water Quality Guidelines (ANZECC & ARMCANZ 2000). Physical measurement parameters include flow, temperature, conductivity, suspended solids, turbidity and colour. These are important parameters themselves, and modify the impacts of chemical stressors. Chemical measurement parameters include pH, alkalinity, hardness, salinity, biochemical oxygen demand, dissolved oxygen and total organic carbon. In addition, other major controls on water chemistry include specific major anions and cations, and nutrient species (phosphate, nitrate, nitrite, ammonia, silica). These controls together with the physical measurement parameters determine the stability, chemical forms and bioavailability of a range of minor and trace contaminants such as metals, metalloids and specific organic compounds. Chapter 5 discusses laboratory approaches to measuring these parameters.

Table 3.4. General measurement parameters used for assessing aquatic system health

Measurement parameter	Input	Potential effects
Electrical conductivity	Salt	Loss of sensitive biota
Total phosphorus	Phosphorus	Eutrophication (nuisance algae)
Ratio of total phosphorus to total nitrogen	Phosphorus and nitrogen	Cyanobacterial blooms
Biochemical oxygen demand	Carbon in organic material	Asphyxiation of respiring organisms, e.g. fish kills
Turbidity	Sediment	Changes in ecosystem habitat
		Loss of sensitive species
		Altered light climate that affects productivity and predator–prey relationships
Suspended solids	Sediment	Changes in ecosystem habitat, loss of sensitive species
Chlorophyll	Nutrients	Eutrophication
pH	Acid drainage	Loss of sensitive biota
Metals, organic compounds	Toxicants	Loss of sensitive species

The protection of aquatic ecosystems is a specific issue discussed in the Water Quality Guidelines that illustrates the selection of physical and chemical measurement parameters. Table 3.4 lists physical and chemical measurement parameters that are frequently used for assessing the general health of aquatic environments. (The term 'health' refers to the condition and functioning of an ecosystem in comparison to conditions and function that are thought to be natural.) Other physical measurement parameters, such as rainfall, catchment morphology, geology, water colour, inflow rates and temperature, may also be crucial causal factors underlying the general measurement parameters.

There have been attempts to integrate various water quality indicators into single indices of water quality, to simplify the presentation and communication of results (Ellis 1989 section 10.2; Ward et al. 1990 pp.73–74; and references therein). Suitable integration of physical and chemical indices can result from the deliberations of water quality experts supported by use of appropriate statistical methods. Another approach is to use multivariate statistical methods (Ellis 1989). However, such integration of indicators reduces the information presented and is not a substitute for detailed presentation of data. There is also the risk that deterioration in one indicator can be masked by improvement in another. Ellis (1989) recommends building up practical experience of the performance of an index over a trial period. Use of indices seems to have been limited in Australia.

3.5.2. Ecotoxicological Assessment

Ecotoxicological studies assess the chronic and acute toxic effects of contaminants on biota in waters and sediments. The studies include the application of laboratory bioassays, and the measurement of biomarkers, focusing on effects at the species level. The ways in which organisms deal with contaminants in terms of bioaccumulation, bioconcentration and regulation are important in determining the ultimate toxic impacts. Ecotoxicological assessment techniques are summarised in Table 3.5.

3.5.2.1. Toxicity Testing Using Sensitive Bioassays

For a limited suite of individual chemicals, it is well recognised that the application of water quality and sediment quality criteria does not necessarily provide adequate protection for aquatic life. Discharges from point sources, such as sewage effluents, and diffuse sources, such as urban runoff, are complex mixtures that contain many unknown compounds which may act together to increase or ameliorate the toxic effect.

Rather than attempting to identify all the chemicals in a sample, toxicity tests (bioassays) using living organisms, are useful as measurement parameters of water quality — in particular the potential toxicity of contaminants in aquatic systems. Bioassays with bacteria, algae, invertebrates and fish are widely used to assess the environmental impact of chemicals in marine and freshwaters, and sediments. Laboratory toxicity data have been used in ecological risk assessments, to derive water and sediment quality guidelines, to investigate the bioavailability of contaminants and to establish cause–effect relationships for particular chemicals.

Acute (short-term) toxicity tests typically measure organism survival over 96 hours or a sub-lethal effect such as light inhibition in light-producing bacteria. Chronic tests determine toxicity over a significant portion of an organism's life span (e.g. weeks, months or years), or use sensitive early life stages such as larvae. Such tests measure, for example, the inhibition of growth rate in microalgae, the inhibition of fertilisation in macroalgae, and developmental abnormalities in scallop larvae. Because different organisms have different sensitivities to the same chemical, batteries of toxicity tests on sensitive species from different trophic levels are currently used. Appropriate test species and ecologically relevant endpoints are selected to suit the aims of the particular study. As toxicity testing is a particularly specialised field of work, advice from experts in this area should generally be sought before opting for particular toxicity tests.

Table 3.5. Summary of ecotoxicological approaches and measurement parameters

Approach	Measurement parameters	Advantages	Disadvantages	Overall value
Single species aquatic bioassays	Algae, bacteria, invertebrates and fish	Ultimate measure of chronic and acute biological impacts which, in combination with Toxicity Identification and Evaluation protocols, can identify and target the source of toxicity	Difficult to extrapolate from laboratory bioassays and mesocosms to the ecosystem as a whole Tend to concentrate on acute not chronic tests; short life-span chronic test species may not be representative	Ultimate measure of water quality with respect to toxicants
Whole sediment or sediment pore water bioassays	Algae, invertebrates	Pore water tests use water organisms; whole sediment tests are better	Difficulty in maintaining field chemistry (redox conditions) in laboratory studies; pore water tests not always on ecologically relevant species	Ultimate measure of sediment toxicity
Biomarker studies	Algae, bacteria, invertebrates and fish (e.g enzyme changes, scope for growth, deformities)	Indicators of chronic stress or exposure	Difficult to relate some changes to specific chemical exposure or to extrapolate from biomarker changes to whole organism or ecosystem effects	Currently more indicative of exposure than effects
Bioaccumulation and biomagnification of toxicants	Principally macro-invertebrates and fish	Some techniques can target particular toxicants, others are non-specific; diagnostic potential good; indication of accumulation of bioavailable chemical contaminants	Require sophisticated equipment for analysis of toxicants, and high level of expertise Need to establish factors affecting bioaccumulation, e.g. size, sex, age, exposure history Difficult to interpret ecological significance	Greatest potential is for detecting known toxicants Can assess exposure to chemical contaminants, but need a very good understanding of intrinsic and extrinsic factors affecting accumulation

Test species can be cultured in the laboratory (phytoplankton, cladocerans) or collected from the field immediately prior to testing, e.g. sea urchins. Tests using readily cultured species have the advantage that they are highly reproducible and are not subject to seasonal availability. Microbial tests, in particular, are rapid, relatively inexpensive, sensitive to some contaminants, and do not have animal ethics constraints on their use. It is important, however, to use a suite of toxicity tests from different trophic levels (OECD 1984), if attempting to relate the results to potential effects in the environment.

Standard toxicity test protocols have been published by OECD (1984), ISO (1989), USEPA (1993, 1994a,b) and Environment Canada (1990a, 1992). The protocols have been adapted to local species of marine and freshwater algae (Stauber et al. 1994; Gunthorpe et al. 1997), invertebrates (Julli et al. 1990; Krassoi et al. 1996; Simon and Laginestra 1997) and fish (Munday et al. 1991; Hyne and Wilson 1997). Choices of test species (marine, freshwater, tropical, temperate) and test endpoints (e.g. survival, growth inhibition, reproduction, developmental abnormalities) depend on site-specific requirements.

For freshwater testing, growth and enzyme inhibition bioassays have been developed with Australian isolates of freshwater green algae such as *Chlorella* species (Stauber 1995; Franklin et al. 2000). Protocols have also been devised with a number of freshwater cladoceran species including *Daphnia carinata*, *Ceriodaphnia dubia* and *Moinodaphnia macleayi* (Julli et al. 1990; Hyne et al. 1996) and the freshwater shrimp *Paratya australiensis* (Abdullah et al. 1994). Acute toxicity tests have also been developed with eastern rainbow fish (Kumar and Chapman 1998) and the tropical purple spotted gudgeon (Markich and Camilleri 1997).

Marine toxicity testing protocols with Australian algal species include growth inhibition and enzyme inhibition tests with diatoms (*Nitzschia closterium*) and green microalgae (Stauber et al. 1994), and fertilisation, germination and growth inhibition tests using macroalgae such as *Hormosira banksii* (Gunthorpe et al. 1995; Kevekordes and Clayton 1996), *Phyllospora comosa* and *Macrocystis angustifolia* (Burridge et al. 1995, 1996). Marine invertebrate tests include inhibition of scallop larval development (Krassoi et al. 1996), fertilisation inhibition in sea urchins (Simon and Laginestra 1997), and amphipod survival (Burridge et al. 1995). Marine fish tests using larval survival of Australian bass and estuarine perch (Hyne and Wilson 1997) have also been developed; however, their application is currently limited by animal ethics restrictions on fish testing, particularly in NSW.

For sediment toxicity testing, an acute 10-day sediment toxicity test and a 14-day chronic test with the estuarine amphipod *Corophium* sp. has been devised (Hyne and Everett 1998), as well as a test using microphytobenthos such as the diatom *Entomoneis* cf. *punctulata* (Adams 2000). An acute 10-day freshwater sediment test has also been developed using nymphs of the mayfly *Jappa kutera* (Leonard et al. in press). While toxicity tests are useful measurements of water quality, they cannot identify specific toxicants causing an observed effect. Toxicity identification and evaluation (TIE) combines chemical manipulation, analytical chemical techniques and concurrent toxicity testing to determine the components of the effluent, water or sediment that are causing the observed toxicity. Protocols for freshwater and marine toxicity identification evaluation have been developed by the USEPA (1991a, 1996a) and modified for use with Australian species in freshwaters (Pablo et al. 1997) and marine waters (Hogan 1998; Doyle 1998).

3.5.2.2. Measurement of Biomarkers

A biomarker is a 'variation in cellular or biochemical components or processes, structure or functions that is measurable in a biological system or sample' (National Research Council 1987). Biomarkers include change in enzyme activity, biochemical changes, physiological changes, histopathological changes and physical deformities. The most intensively studied group of organisms is fish, particularly marine fish (Holdway et al. 1995). A number of studies have examined the abilities of sub-cellular biomarkers to respond effectively to, and generally indicate, the effects of a number of chemical contaminants. The most promising use of biomarkers is as a screening tool for the detection of exposure to contaminants, e.g. with mixed function oxidases. However, biomarkers give little information about the effects of chemical contaminants and it is difficult to relate biomarker changes to effects at the individual organism, population, community or ecosystem level.

Biomarkers have been applied worldwide in both freshwater and marine ecosystems. Applications include investigation of deformities in chironomid mouthparts (Warwick 1989), and cellular and enzymatic changes (Gunther et al. 1997; Viarengo et al. 1997). At this stage in Australia, biomarkers have been developed only for estuarine and marine systems, e.g. biomarkers in flathead to detect contaminant exposure (Holdway et al. 1994, 1995), histopathological changes in flounder (Stauber et

al. 1996) and other organisms (Twining and Nowak 1996) and enzymatic changes in algae (Peterson and Stauber 1996).

3.5.2.3. Measurement of Bioaccumulation

Two major difficulties in identifying toxicant impacts on ecosystems are that contaminant events may be episodic and that toxicants may have effects at very low concentrations. Where a particular toxicant is known or suspected to occur, monitoring of toxicant levels in biota can be a useful technique, particularly when concentrations of toxicants in water are too low to be measured chemically.

There are two biological phenomena that can assist with this approach: bioaccumulation and biomagnification (Connell 1981). Bioaccumulation refers to the continued accumulation of particular contaminants by some taxa throughout their lifetime, e.g. accumulation of trace metals by molluscs. The existence of this phenomenon means that episodic contamination events will be integrated over time in an organism's tissues. Biomagnification refers to the increase in concentration of a contaminant up the food chain, such as occurs with some organochlorine pesticides. This phenomenon makes it easier to detect low concentrations of chemical contaminants in an ecosystem. A disadvantage of this approach is that the toxicant must be known. With the increasing complexity of industrial and other effluents, it is not always possible to identify the key toxicants.

3.5.2.4. Early Detection of Change

Sub-lethal tests can be part of programs aiming for early detection of change. If potential adverse effects of a disturbance to an ecosystem can be predicted or detected quickly, more substantial and damaging effects may be avoidable through management action. Specific and sensitive early detection programs may be set up to provide, firstly, predictive information from laboratory-based direct toxicity assessment, and secondly, early detection in the field. Each involves measurement of sub-lethal responses by organisms.

Measurements of bioaccumulation, or biomarkers, or bioassays using sensitive species, as discussed above, all are techniques that might be used. They provide different information. Bioaccumulation is an organism-specific integration of water (or sediment) contaminants, so particular compounds must be measured. Biomarkers and bioassays provide more generic responses to acute or chronic impacts; chemistry then identifies the specific stressor that requires management. For more information on bioassays refer to the Water Quality Guidelines section 8.3.6, on direct toxicity assessment (ANZECC & ARMCANZ 2000 Volume 2).

3.5.3. Ecological Assessment

Ecological assessment aims primarily to measure the structure and function of biological communities. It principally involves field-based measurements that examine effects on the relative abundance and diversity of species, community structure and composition, and how these are altered as a consequence of known or unknown stressors and their modifiers in both waters and sediments. A summary of techniques for ecological assessment is given in Table 3.6, which also shows the variety of taxonomic groups that have been tried out for ecological assessment of ecosystem health.

Macroinvertebrates have been selected as the key indicator group for bioassessment of the health of Australia's streams and rivers under the National River Health Program (Schofield and Davies 1996). Active monitoring programs by state and territory agencies using the AUSRIVAS–RIVPACS-type approach (see Box 1, page 3-24), and underpinned by extensive R&D, are well established across the country. A large number of studies have been published on the use of benthic invertebrates in the assessment of both freshwater and estuarine water quality in Australia (e.g. *Australian Journal of Ecology* volume 20, number 1, 1995; Deeley and Paling 1999). The responses of various invertebrate taxa to specific types of chemical contamination are reasonably well documented for Northern Hemisphere waters (e.g. Hellawell 1986).

Box 1. AUSRIVAS — what it is, how it works and its origins

The Australian River Assessment System (AUSRIVAS), and its basis, the British RIVPACS system (River Invertebrate Prediction and Classification Scheme), are rapid standard methods for rating the ecological health of freshwaters by biological monitoring and habitat assessment. Both AUSRIVAS and RIVPACS consist of models that predict the fauna (usually macroinvertebrates) expected to occur at a test site on the basis of its environmental attributes — its geographic, physical and chemical features. When the test site is sampled, the fauna observed are compared to the models' expectations for that sort of habitat, and the resulting observed/expected (O/E) score is an integrated indicator of river health.

AUSRIVAS was developed over the period 1993–1997, and has established protocols for invertebrate sampling, habitat assessment and model development. It is now being used in a national assessment of river health involving some 6000 sites to be sampled over three years. In the development of AUSRIVAS, aquatic invertebrates at more than 1500 minimally disturbed sites (reference sites) were sampled across Australia to establish a reference site database from which to build the predictive models. Sites were sampled in two seasons: autumn and spring for temperate regions, and the wet and dry seasons in the tropics. Five types of habitat were sampled and the two most common aquatic habitats from riffle, edge, main channel, macrophytes and pool rocks were selected for use in constructing a model in each state or territory. Macroinvertebrates were chosen because they have these valuable measurement parameter characteristics:

- the fauna is well known taxonomically;
- the fauna is diverse with known differences in response to different contaminants;
- there are enough individuals within a sample to provide abundance data to be used in analysis, and yet numbers are not unmanageable;
- generation times of taxa are such that they integrate ecological impacts over a satisfactory period.

The AUSRIVAS O/E score is responsive to a variety of environmental effects, including water quality, habitat condition, and changes in flow regime. Various O/E score categories (bands) are used to provide a 'biological thermometer' of the overall condition and severity of disturbance for various sites. This allows the general health of the waters at the survey sites to be characterised in a nation-wide context. The AUSRIVAS scores do not provide a clear indication of the cause of a disturbance or contamination in waterbodies, but they can be used be used to identify those waterbodies that are 'stressed' or those that need further investigation and management action. The method depends on the existence of reference sites that are close environmental equivalents to the test site. If a test site's environmental attributes do not match any in the reference set, its health cannot be assessed unless suitable new reference sites are added to the database.

For more information, see http://ausrivas.canberra.edu.au/ (Coysh et al. 2000), Schofield and Davies (1996), Kay et al. (1999), Marchant et al. (1999), Turak et al. (1999).

Table 3.6. Summary of ecological assessment approaches and measurement parameters

Approach	Measurement parameters	Advantages	Disadvantages	Overall value
Diversity indices	Various	Provide summary of complex data; easy to understand, allow comparisons between sites or times	Ecological significance of indices is unclear; can be affected by sampling and analytical factors	Attractive for their simplicity, but their ecological value is questionable
Biotic indices	Principally macro-invertebrates and algae	Simple, easy to interpret summaries of complex data; can provide contaminant-specific response	Detailed knowledge of contaminant tolerance required for diagnostic use	Usefulness limited by baseline; site-specific and contaminant tolerance information needed
Stream community metabolism	Benthic flora and fauna	Integrates impact across the entire benthic biota; relatively rapid; provides a simple output	Technique not proved; may be less useful in disturbed catchments; diagnostic capability unclear	Technique has potential, but its sensitivity and diagnostic capacity have not been demonstrated
Macro-invertebrate community structure (e.g. AUSRIVAS) for rapid biological assessment; quantitative methods for site-specific studies	Macro-invertebrates	Integrates over appropriate temporal and spatial scales; much background information available; good diagnostic capability	Relies on complex modelling approach; output not as readily understood as other techniques	Great potential for identification of impacts; reasonable potential for establishing causes of impacts
Macrophyte community structure	Macrophytes	Easily sampled, respond to a range of impacts	Gives poor under-standing of factors affecting community structure; insensitive to some chemical contaminants	Limited use
Fish community structure, biomarkers (biochemical, physiological, immunological or histo-pathological)	Fish	Readily sampled, taxonomically well known	Gives poor knowledge of population dynamics and water quality factors; temperate fauna are impoverished; biomarker techniques require sophisticated equipment and high level of expertise	Community structure uses more applicable in tropical than temperate waters
Algae: biomass and community structure	Algae	Sensitive, taxonomically well known, has diagnostic potential; community structure (AUSRIVAS-type) approach most promising	Identification requires high level of expertise; community structure approach not well tested	Community structure approach has good potential
Bacteria, protozoa and fungi: community structures	Bacteria, protozoa and fungi	Organisms occupy key ecological role so community change can provide valuable key to impacts	May recover too rapidly from impact for monitoring purposes; taxonomy and response to chemical contaminants poorly known	Limited use at present; would require extensive taxonomic and diagnostic work before they could be useful

Fish have considerable potential for use in the bioassessment of water quality in some locations (Harris 1995). Australia has a freshwater fish fauna that is highly diverse in the northern part of the continent but of low diversity in southern and inland regions. Fish populations and communities respond to changes in water quality but are also strongly influenced by changes in hydrology (which affect recruitment, habitat and food availability) and physical habitat structure (such as organic debris, bottom substrate and pool dimensions, and barriers to migration). Fish have been little used for assessing water quality or human-induced effects on water quality except in the bioassessment program at Ranger uranium mine (Humphrey et al. 1990) and in a series of studies in south-east Queensland (summarised in Arthington et al. 1998). Current attempts to develop standardised bioassessment approaches using fish are in their infancy in Australia. None of these methods is applicable at a broad scale and none has been extensively tested to date. Approaches based on comparative measures of community composition are compromised in most of southern and inland Australia where species diversity is low, fluctuations in species abundance and occurrence are extreme (driven by unpredictable flow events), and the relative dominance of exotic species is high. Together, these factors mean that fish diversity can be strongly affected by the chance appearance or disappearance of a single species, and so it is often a poor measure of ecological integrity, requiring further work. Fish populations and communities change and respond to environmental factors at much longer time scales than most other aquatic biota (e.g. algae, macroinvertebrates).

Periphytic diatoms have been used in streams and rivers in the United Kingdom (Whitton and Kelly 1995) and, more recently, in Western Australia, South Australia, Victoria and New South Wales as part of the National River Health Program (John 1998). The Western Australian and New South Wales–Victorian work has demonstrated that ecologically disturbed and undisturbed sites consistently have different assemblages of diatoms, suggesting that this approach could be used routinely to identify disturbance. This work has also found that certain groups of diatoms are well correlated with water contaminants; i.e. the technique may have diagnostic potential.

Bacteria, protozoa and fungi have not been widely used in ecosystem health studies, but bacteria and protozoa have been used extensively to test that waters are safe for human use.

The potential for the development of macrophyte bioassessment procedures for Australian streams is currently being evaluated under the National River Health Program, but at the moment there are few universally applicable protocols available.

Before choosing a particular taxonomic group as a measurement parameter of water quality or ecosystem health, the monitoring team should check that the taxonomic group fulfils this set of criteria:

- the measured response reflects the ecological condition or integrity of the site, catchment or region to be monitored;
- approaches to sampling and data analysis can be highly standardised;
- the response can be measured rapidly, cheaply and reliably;
- the response has some diagnostic value.

3.5.3.1. Measures of Macroinvertebrate Community Structure

Macroinvertebrate communities provide the most developed indication of ecological health. Invertebrate data are analysed by aggregating them into measures or indices. For example, the Macroinvertebrate Community Index (MCI) (Stark 1985, 1993) has been developed in New Zealand and is now widely used there by regional councils to detect and monitor water quality degradation. Similarly, Chessman (1995) has developed the SIGNAL index (Stream Invertebrate Grade Number — Average Level) for invertebrates identified to family level in south-eastern Australia. These measures are forms of biotic indices. They are based on the premise that contaminant tolerance varies between species or higher taxa; they produce contaminant tolerance scores.

Two problems arise in developing and applying contaminant tolerance scores in Australia and New Zealand. First, most tolerance information relates to organic contaminants, although knowledge of

tolerances to acidification and heavy metals is growing; also, in Australia, most information on contaminant tolerances comes from the wetter and better studied temperate south-east and south-west of the country. Second, many tolerance indices are developed for family or coarser-level identifications (many groups of invertebrates are hardly known, taxonomically, beyond this level), but it is acknowledged that for some groups the constituent taxa may vary widely in their tolerances.

Four generic biodiversity-type protocols and four early-detection-type protocols have been developed for streams and wetlands using macroinvertebrate species or communities (see the Water Quality Guidelines Table 3.2.2 and Section 8.1.3).

Another biotic index, the ETP index, is based on Ephemeroptera, Trichoptera and Plecoptera. These three families of macroinvertebrates are sensitive to most types of contaminant, and so the numbers of individuals in these orders should decrease with a decrease in water quality. The numbers of some Diptera and tubificid worms may increase in response to contaminants, especially organic contaminants. This response has been used in constructing indices by examining the ratios between tubificids and other organisms, or by just counting the numbers of taxa belonging to the sensitive groups (e.g. Plafkin et al. 1989). Virtually all the indices or other measurements using these assumptions have been developed in relation to organic contaminants in rivers. However, some species of Trichoptera and Ephemeroptera are highly tolerant of trace metal contaminants (Norris 1986), so caution is advised in the general application of indices based on the assumptions just discussed. Other difficulties include the large amount of initial work that may be needed to define contaminant tolerances and to define 'clean' freshwater communities, and the limited number of taxonomic keys to many species (Resh and Jackson 1993).

Waterwatch programs across Australia collect macroinvertebrate samples as part of regular spring and autumn surveys. Using a national protocol, Waterwatch collects data on the abundance and diversity of macroinvertebrates identified to the order level only, at regular sites in local waterways. At this stage no adequate diagnostic tool is available to convert these data into a biotic index, though various trials have been conducted. Waterwatch is working with scientists to develop a modified SIGNAL-type scoring system for this purpose, and the data are being collected and compiled throughout Australia. Aside from the scientific purpose of collecting data, the collection and processing of macroinvertebrate samples by the general community has enormous educational value. Further information can be found on the Waterwatch web site (www.waterwatch.org.au).

The SIGNAL score is widely used and has been shown repeatedly to have strong relationships with water quality variables (Growns et al. 1995, 1997; Chessman et al. 1997; Chessman and McEvoy 1998). Research is currently being conducted through the National River Health Program to develop and test a series of other indices for macroinvertebrate community structure to assess river health and the impact of water quality changes.

The most widely used protocol is the Australian River Assessment System, AUSRIVAS, based on the RIVPACS model developed in the UK (see Box 1, page 3-24). Through the National River Health Program, the first Australia-wide assessment of the health of Australia's diverse and unique aquatic systems has been undertaken using macroinvertebrates and the AUSRIVAS method.

3.5.3.2. Rapid Biological Assessment

The Water Quality Guidelines make prominent reference to rapid biological assessment (RBA) in the context of rapid and cost-effective techniques for obtaining first-pass, not necessarily quantitative, data over broad geographical areas. Rapid techniques are suitable for determining the extent of a problem such as river health (see the Water Quality Guidelines section 3.2.1.3 and section 8.1.1.1). Further discussion on RBA approaches using stream macroinvertebrate communities is provided in Resh and Jackson (1993), Lenat and Barbour (1994) and Resh et al. (1995).

The most commonly used RBA method in Australia is AUSRIVAS. Rapid bioassessment protocols are also being developed for riverine benthic algae (diatoms) and fish, as well as for

macroinvertebrate communities in wetlands and estuarine sediments (see the Water Quality Guidelines section 3.2.1.3/1).

The data obtained from RBA may be suitable for broad-scale auditing or screening purposes and for broad-scale management and for use in an early warning system. In most cases RBA should be followed by detailed studies using quantitative methods for site-specific assessments.

The Water Quality Guidelines advises against the use of RBA methods alone for detailed site specific assessments and comparisons in time and space, though these methods may usefully support quantitative methods in this role (see Water Quality Guidelines Table 3.2.1, Table 8.1.2 and section 7.2.1.1/1). Rapid biological assessment methods have been suggested for monitoring the outcome of environmental flow studies or river rehabilitation or river restoration projects and for setting baselines for environmental impact assessment. These issues are complex because the RBA approach compromises various aspects of sampling design and its implementation, and that can have cumulative effects on the results obtained. For example RBA methods use sample collection methods that are simple in comparison to those required for quantitative assessment. Likewise, RBA may sample only single test sites rather than the set of sites required for statistical purposes to account for between-site variation. Also only a sub-sample of the animals collected may be identified, and only to the family or genus level rather than to species level. Abundance data for the taxon identified are not used by the models. All these elements when combined tend to make it difficult to 'gear-up' the RBA methods for quantitative analysis. For example, working out how many AUSRIVAS samples would need to be collected upstream and downstream of a point source of contaminants to monitor their impact and the selection of appropriate control sites may be irrelevant. The limitations of the other aspects of the protocol such as the collection methods, sample processing and analyses methods may not provide the level of detail required for coping with the variation between samples. The cumulative effect of these limitations restricts RBA methods, making it difficult to apply them or modify them for uses beyond those for which they were designed. The suggested appropriate application of RBA and quantitative methods is summarised in the Water Quality Guidelines Table 8.1.2 and the differences in the protocols for these methods are given in the Water Quality Guidelines Appendix 3 Methods 3A(i) and 3A(ii).

3.5.3.3. Whole Ecosystem Ecological Assessments

The conservation, maintenance, rehabilitation and restoration of healthy aquatic ecosystems and biotic integrity have become important objectives of water management worldwide (Gore 1985; Karr 1991; Rapport 1991) and also in Australia (Norris and Thoms 1999). The focus on healthy ecosystems also applies to lakes, wetlands, estuaries and other water bodies including groundwater ecosystems.

The term 'health' is usually defined in terms of ecological integrity (Schofield and Davies 1996; Karr and Dudley 1981) as:

> the ability of the aquatic ecosystem to support and maintain key ecological processes and a community of organisms with a species composition, diversity, and functional organisation as comparable as possible to that of natural habitats within a region.

Several single integrated measures of the integrity or health of aquatic ecosystems have been developed. Principally these are the *diversity indices* and *biotic indices* and the application of various biological measurement parameters, either on their own or combined with other measures for monitoring water quality. It is generally agreed that an integrated approach is desirable, but in practice several agencies are monitoring single measures as a priority.

Another approach makes integrated measurements of the metabolism of an aquatic community or some other ecosystem process involving nutrients, production or carbon metabolism. Ecotoxicological methods have also been used. Gower (1987) and Legendre and Legendre (1998) give guidance on the use of integrated techniques in freshwater. The Land Ocean Interactions in the

Coastal Zone Program (LOICZ) has considered community metabolism and modelling approaches to marine biodiversity; see http://kellia.nioz.nl/loicz/

3.5.3.4. Diversity Indices

Diversity indices usually require a count of the total number of individuals and a total count for each of the taxa. The taxa need to be separated but not necessarily identified. Separation is often at the species level but it is sometimes at the generic or family level. A higher diversity, i.e. the presence of more taxa within a given number of animals, is taken to signify a healthier ecosystem. A disadvantage is that measurements require taxonomic skills, are tedious and require large numbers of samples to achieve statistical significance.

Despite this, diversity indices have been widely advocated as measurement parameters of ecosystem health for a variety of reasons:

- they are seen as a useful way to condense complex data and thus aid interpretation;
- people with little biological expertise can easily understand them, and can gather the data to create some of them;
- they are a more generic measure than physical and chemical measures;
- they allow comparisons between sites or times where collections have been made using different sample sizes, methods or habitats.

The combination of evenness and taxonomic richness in a diversity index supposedly indicates the state of the community. It seems to be generally accepted that index values decrease with decreasing water quality. Low diversity is taken to indicate a stressed community that tends to be unstable.

In practice, there are several matters for concern with the application of diversity indices. First, statistical anomalies occur with some indices as a result of the assumptions on which they are based. Simpson's index is based on the assumption that in more diverse communities there is a lower probability that individuals chosen at random will belong to the same taxon (Simpson 1949). However, this assumption disregards the possibility that members of the same taxon will be clumped for reasons of microhabitat, breeding, or behaviour.

Second, in the diversity indices of Gleason and Margalef which are similar and are based on guesses at fitting curves to species abundance distributions (Gleason 1922; Margalef 1958), the assumption is made that the number of individuals is directly proportional to the area sampled. Of course, these indices are likely to be highly dependent on sample size. There is some doubt about the biological meaning of frequency distributions, and it is not clear how environmental stress (including contaminants) will affect the relationship.

Third, a problem with the most widely used diversity indices, those derived from information theory, is the rather tenuous biological significance of the measure. For example the Shannon index (Shannon 1948) reaches its maximum value when all species are evenly distributed. Biologically, this is assumed to be the most desirable situation, although it contradicts the evidence provided by the log-normal distribution for many different communities. A range of factors other than contaminants may affect this type of diversity index, including sampling method, sample size, depth of sampling, duration of sampling, time of year, and taxonomic level used. Diversity indices based on information theory should be interpreted and compared with caution.

In summary, the measurement of diversity and the effect of contaminants on a diversity index both must be resolved before these indices can be used and the results interpreted effectively. When such indices are applied, the predicted effects of contaminants on the ecological attributes supposedly measured by the index are rarely stated. This criticism might also be applied to other indices and to most other approaches.

The use of diversity indices in Australia has tended to focus on those groups for which the taxonomy is best known, and which are abundant and taxonomically rich enough to provide reliable measures of diversity. Of particular interest have been aquatic macroinvertebrates. However, the extent of the

work on predictive modelling using macroinvertebrates (such as in AUSRIVAS, in which both sides of the O/E ratio are measures of the composition of the fauna) has overshadowed the use of diversity indices in Australia.

3.5.3.5. Biotic Indices

Biotic indices usually have been developed empirically as a means for assessing the effects of contaminants, mostly in rivers. Many are specific to a site and contaminant type (usually organic). The calculation of biotic indices usually requires:

- a total count of individuals or total counts of taxa;
- counts (or biomass measurements) of specific groups such as all insects and tubificid worms, or the number of mayflies, stoneflies and caddisflies;
- detailed lists of the responses of different taxa to contaminants; or
- division of invertebrates into groups with different feeding strategies.

Biotic indices are more clearly related to the conditions that led to their development than are diversity indices.

Some examples of biotic indices and community indices are these: the Invertebrate Community Index (ICI) (DeShon 1995); the Rapid Bioassessment Protocols used by USEPA (Shackleford 1988; Plafkin et al. 1989; Barbour et al. 1992, 1995, 1996; Hayslip 1993; Smith and Voshell 1997); and various Macroinvertebrate Community Indices (MCI), including quantitative and semi-quantitative applications, used in New Zealand (see also section 3.5.2.4; Stark 1985, 1993, 1998; Collier et al. 1998). The Index of Biotic Integrity (IBI), originally developed for fish (Karr 1981; Karr et al. 1986; Miller et al. 1988), uses species richness, abundance, community structure, and the health of the individual animals as metrics. It has been tried in Australia (Harris 1995; Harris and Silveira 1999), but its potential application is still under investigation. A Benthic macroinvertebrate Index of Biotic Integrity (B-IBI) (Kerans and Karr 1994; Fore et al. 1996) has also been developed.

The metrics used in these indices evaluate aspects of community composition and of structure and processes within the macroinvertebrate assemblages. Although the indices have been developed first for a particular region, they are typically applicable over wide geographic areas with minor modification (Barbour et al. 1995).

The USEPA has developed the Rapid Bioassessment Protocol for Use in Streams and Rivers (Plafkin et al. 1989); it assesses community diversity as a measure of water quality. Contaminants are indicated by the absence of contaminant-sensitive benthic macroinvertebrate groups (Ephemeroptera, Plecoptera, and Trichoptera) and the dominance of contaminant-tolerant groups (oligochaetes or chironomids). Overall, a paucity of benthic macroinvertebrates may indicate impairment. However, nutrient levels are naturally low in pristine headwaters and may explain the low productivity and few benthic macroinvertebrate species that are found there.

The most effective metrics are those that respond across a range of human influence (Fore et al. 1996; Karr and Chu 1999). Resh and Jackson (1993) tested the capacity of 20 benthic metrics used in 30 different assessment protocols to discriminate between impaired and minimally impaired sites in California, USA. The best measures, from their study, were the richness measures, two community indices (Margalef's and Hilsenhoff's family biotic index), and a functional feeding group metric (percent scrapers). Resh and Jackson (1993) emphasised that both the measures (metrics) and the protocols need to be calibrated for different regions of the country, and, perhaps, for different impact types (stressors). In a study of 28 invertebrate metrics, Kerans and Karr (1994) demonstrated significant patterns for 18 metrics and used 13 in their final B-IBI. Richness measures were useful, as were selected trophic and dominance metrics. One of the unique features of the fish IBI (Harris and Silveira 1999) which is not found in benthic indices is that it can incorporate metrics on individual condition. However, Lenat (1993) has advocated measures evaluating chironomid larvae deformities.

3.5.3.6. Similarity Measures

Communities of organisms at two sites can be similar or dissimilar. The similarity (or dissimilarity) can be measured and rated, and related to water quality that has been assessed using other parameters, perhaps physical or chemical.

Numerous numeric similarity (or dissimilarity) indices have been proposed. The PATN package (Belbin 1993; Belbin and McDonald 1993), developed in Australia, includes a wide range of similarity measures and techniques for examining community structure and patterns in species distribution, and association and similarity between sites and species.

3.5.3.7. Functional Feeding Group Measures

Some biotic indices are based on the assumption that the ratios of organisms with different feeding strategies will change with contamination (e.g. collectors will be more abundant than shredders under contaminated conditions), or that trophic generalist organisms will be more tolerant to contaminants than trophic specialist organisms. There is some doubt that these general rules hold true, and even that it is possible to assign taxa to different feeding strategies (Resh and Jackson 1993).

Organisms can be assigned to groups of species that feed in the same way, generally known as functional feeding groups. For example, bacteria can be photosynthetic autotrophs, or bacterivores–detritivores, or algivores, or nonselective omnivores, or saprotrophs, or raptors. Feeding groups have also been identified for fish: they can be predators, grazers, strainers, suckers or parasites. The usefulness of functional feeding groups has not been well demonstrated for benthic macroinvertebrates, and the concept is not considered reliable in this case (Karr and Chu 1997).

Feeding groups represent a way of assessing the dynamics of food supplies in a water body and the balance of feeding strategies (food acquisition and morphology) in the fauna assemblages. An imbalance in functional feeding groups generally indicates stressed conditions. Specialised feeders, such as scrapers, piercers, and shredders, are relatively sensitive organisms and are thought to be well represented in healthy streams. Generalists, such as collectors and filterers, have a broader range of acceptable food materials than specialists (Cummins and Klug 1979), and thus are more tolerant to contaminants that might alter availability of certain food. However, filter feeders are also thought to be sensitive in low-gradient streams (Wallace et al. 1977).

3.5.3.8. Taxonomic Richness

Taxonomic richness generally decreases with decreasing water quality. The number of individuals and biomass may increase or decrease, depending on the type of contaminant and the organisms involved. Considerable work is needed to classify adequately the responses of different taxonomic groups to contaminants, an increasingly difficult task because new chemical contaminants continue to appear, and because many effluents (including sewage) are becoming more complex.

3.5.3.9. Stream Community Metabolism

The stream community metabolism approach is based on the concept that movement of organic carbon through an ecosystem can be used as a measurement parameter of stream community metabolism. This in turn provides an indication of ecosystem health.

Two biological processes affect the movement of carbon: production (via photosynthesis) and respiration. It is argued that community metabolism is sensitive to small changes in water quality (particularly organic contaminants and sedimentation) and to riparian conditions that affect light input. As a result of this sensitivity, the stream community metabolism approach may be able to detect a disturbance early, before it is manifest in changes in organism assemblages (e.g. macroinvertebrate community composition).

The two key community metabolism components measured are gross primary production (P) and respiration (R). The ratio of these measures provides a measurement parameter of stream status and health. Undisturbed forest stream sites are typically heterotrophic (e.g. $P/R<1$) and so are net

consumers of carbon. Sites in which the P/R ratio >1 (i.e. fundamentally autotrophic ecosystems) have been found either in cleared catchments or in sites that have been nutrient enriched. Consequently, the P/R ratio is an index with ecological meaning in freshwaters.

Community metabolism is best measured by monitoring water oxygen concentration. In systems of high or rapid metabolism, whole-river measurements can be made using the two-station (e.g. Odum 1956) or single-station technique over 24 hours (e.g. Bunn et al. 1997). In systems with low metabolic rates or high re-aeration due to turbulence (e.g. forested upland streams), closed system procedures are recommended (e.g. Davies 1997). These can be conducted over 24 hours (Davies 1997) or over short time periods in full sunlight followed by no light (Hickey 1988).

The approach uses Perspex chambers placed over the selected habitat and pushed into the substrate. Chambers require an oxygen sensor and a recirculation pump to maintain a flow similar to that outside the chamber. Oxygen concentration is recorded over an appropriate period to allow an accurate calculation of both P and R. A specific habitat can be chosen as a measurement parameter of stream health; then changes in community metabolism are more likely to result from ecological impact than from differences in habitat. Bunn et al. (1999) recommend the use of cobble habitat as the one in which variation in community metabolism best reflects catchment characteristics. Alternatively, the range of habitats present in the stream can be monitored and, using areal weighting, the metabolism of an entire reach can be measured.

The stream community metabolism approach has been applied in Jarrah forests of south-west Western Australia (WA), and in the Johnstone River in north Queensland and in the Mary River catchment in south-east Queensland.

Bunn et al. (1999) contend that gross primary productivity (GPP) could be used to infer the type of impact occurring; increased GPP indicating nutrient enrichment and catchment clearing, and depressed GPP indicating sedimentation or degradation in water quality. Initial results from the Mary River catchment suggest clear links between community metabolism and contaminants or catchment condition. There appears to be potential for development of stream community metabolism as a measurement parameter of stream ecosystem health

3.5.3.10. Quantitative Ecological Assessment

A ‘quantitative’ method refers to one that permits rigorous and fair tests of the potential impacts under consideration; typically, conventional statistical tools are employed to attach formal probability statements to the observations; see the Water Quality Guidelines section 3.2.1.3 and section 8.1.1.3.

The rationale for using quantitative methods is that they allow the use of sampling designs based on statistical inference and, hence, the explicit identification of effect size and Type I and Type II error rates. These procedures are likely to be more sensitive to subtle impacts than those based on rapid bioassessment techniques, where the effect size and error rates are implicit in the modelling procedure. In addition, quantitative procedures based on statistical designs can be adapted to local site-specific conditions.

Where possible, paired areas should be employed in an MBACI design. The current lack of knowledge about year-to-year variations in benthic diversity in many ecosystems argue for at least three years of pre-impact baseline data wherever this is possible. Quantitative sampling methods should also be used. Further details on a protocol for quantitative assessment for macroinvertebrates is provided in the Water Quality Guidelines Appendix 3, method 3A(ii).

This protocol also provides a model for development of protocols for biological indicators such as fish, macrophytes, diatoms and other groups. The key aspects are the survey design and the sampling required to obtain the statistical power needed to produce the desired results or sensitivity in the monitoring program.

3.5.3.11. Selecting Ecological Assessment Methods

If the monitoring team decides to use ecological measurement parameters, that decision will dictate the monitoring strategy and approach required and the various protocols. The team will need to check that the objectives of the monitoring program are not jeopardised. For example, rapid biological assessment methods such as that used for AUSRIVAS may not meet the quantitative assessment requirements of a site-specific study of impact assessment. This applies to the collection, processing and analysis of samples, and to the sampling design needed to meet the statistical requirements of the study. These conflicts are discussed more fully in the Water Quality Guidelines (ANZECC & ARMCANZ 2000) which explicitly distinguishes between rapid biological assessment (RBA) and quantitative analysis, and provides different protocols for these broad types of assessment.

The choice of the right method is a crucial part of the study design. Similarly the monitoring team must choose the most appropriate approaches for the problems or issues under investigation: they could be whole ecosystem approaches, river health assessments, biodiversity indices, community indices or specific indicator species or taxa for particular water quality problems. The monitoring designs that incorporate these measures will be different in each case.

3.6. Data Requirements

Once the decisions have been made about the study type, study boundaries and measurement parameters, the data requirements need to be summarised. The data requirements include the measurement parameters, scale, geographic locations and length of study, frequency, accuracy and precision. These serve as the 'concrete' instructions for the decisions that have to be made about techniques required for data analysis (Chapter 6) and for the design of specifically tailored sampling and analysis programs (Chapters 4 and 5).

3.7. Cost-Effectiveness of Sampling Programs

It is preferable for the cost of sampling programs to be as small as possible while still meeting the stated objectives of the monitoring study. Cost-effectiveness considerations involve trade-offs between loss of statistical power for discriminating between various hypotheses and the cost of data acquisition. It is necessary to determine all the resources and associated costs required, thereby ensuring the study can be carried out. Costs of data acquisition are determined by:

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

There is extensive information available about the optimisation of sampling programs with regard to precision and cost (Montgomery and Hart 1974; Eberhart 1976; Ellis and Lacy 1980; Short 1980; Bailey et al. 1984; Lettenmaier et al. 1984; Hayes et al. 1985; Radford and West 1986; Kratochvil 1987).

3.8. Reporting Schedules

During the study design process it is important that the primary users and the suppliers of the information agree on the reporting schedules. If the expected schedules are unreasonable,

compromise arrangements need to be made. Promising more than can be delivered within a certain time places unnecessary pressure on those doing the monitoring study, while failure to report findings on time will damage relationships between the information user and the supplier.

All stages of the monitoring program will have their own time frames that must be considered when agreeing to a reporting schedule. The monitoring of a range of river flows, for example, will take months or years; with laboratory analyses, the time frames for reporting will vary significantly depending on the analyte.

The design process should consider also the reporting needs and expectations of all other stakeholders and information users. How these might be addressed is discussed in more detail in Chaper 7.

Chapter Four
Field Sampling Program

4.1. Introduction

The study design phase, described in Chapter 3, broadly specified the measurement parameters that were needed for satisfying the monitoring program objectives. Now that the basic outline of a sampling program has been settled, the next stage is the implementation of this design in the field.

First, the monitoring team defines or specifies the population that is to be sampled. Then it considers the specific data requirements — measurement parameters, scale and frequency of sampling, accuracy and precision required — and decides whether to measure the parameters in the field or the laboratory. Costs must be planned so that they fall within the agreed budget, remembering the trade-off between maximum statistical power and cost of sampling and analysis.

In all instances, there are appropriate protocols for field measurements, and for sample collection, preservation, preparation and storage, that need to be followed prior to any laboratory analyses. A framework for applying these protocols is shown in Figure 4.1. A checklist is presented in Table 4.1.

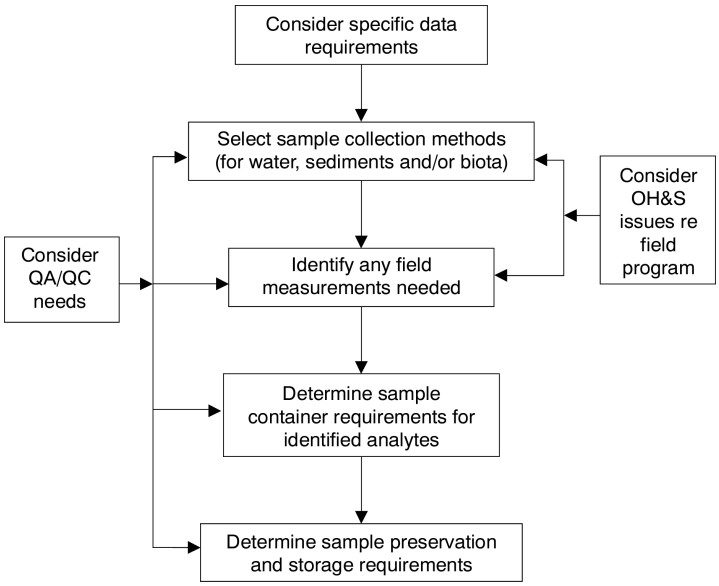


Figure 4.1. A framework for designing sampling programs

4.2. Field Measurements and Observations

Some parameters (e.g. flow, temperature) can only be measured in the field. For other parameters (e.g. dissolved oxygen, redox potential and possibly pH), field measurements are highly desirable because the value of the parameter might change in the sample after collection. The reliable sensors

that are available nowadays make it often convenient to measure many parameters in the field. Not only does field monitoring give on-the-spot values, but also the results can be checked immediately and then the choice of sampling sites can be refined rapidly if necessary. Whether measuring or sampling, quality control and quality assurance are important; they require planning because they are not easy to achieve in the field.

Table 4.1. Checklist for designing sampling programs

1. Have the specific measurement parameters been identified and the data requirements stated?
2. Can the data required be obtained by field measurements?
3. Have appropriate field measurement techniques, including calibration procedures, been selected?
4. How are the positions of sampling sites to be recorded?
5. What ancillary field observations are to be taken?
6. Will the sampling device collect a representative sample?
   (a) Do disturbances occur in the environment being sampled?
   (b) Will the sample be altered by contact with the sampling device?
   (c) Will the sample device contaminate the sample? If yes, how is the sample device to be cleaned?
   (d) What are the effects of the sampling device being in contact with media other than the sample of interest?
7. How are samples to be collected to prevent contamination?
8. Will the sample container contaminate or affect the stability of the sample? If so, how are these problems to be overcome?
9. What size sample containers are required?
10. How are samples to be preserved before analysis?
11. Are procedures in place to track samples and field data?
12. What program is in place to identify, measure and control errors?
    (a) Have sampling protocols been written?
    (b) How are sampling staff to be trained?
    (c) How is the sampling staff's competence to be tested?
    (d) Can the integrity of the sample be guaranteed?
    (e) Have blanks, duplicates and replicates been incorporated into protocols?
    (f) How are problems to be rectified?
13. Are there enough resources to prevent any bottlenecks occurring in field or laboratory that would hinder analyses and compromise data quality?
14. How are data to be stored and accessed?
15. Have all reasonable steps been taken to protect health and safety of employees?
    (a) Have possible hazards been identified and documented?
    (b) Have sampling staff been made aware of possible hazards, and have risk minimisation plans been developed?
    (c) Have sampling staff been trained to ensure that sampling is done safely?
    (d) Will sampling staff be appropriately supervised during sampling activities?

Field data can also be obtained automatically and by remote sensing, and the data can be logged and/or transferred to laboratories by telemetry. This has the advantage of providing measurements that are either continuous or at fixed intervals, allowing very cost-effective studies of temporal trends.

For parameters that do not change during transport and storage, field sampling is adequate. Macro-invertebrates, for instance, collected by sampling a waterbody, can immediately be stored in alcohol in vials and kept there until they are identified and counted. However, for samples that must be field-

sampled and then analysed in a laboratory, fixative, preservatives and cold storage during transport can minimise changes.

Guidance on the various physical, chemical and biological measurement parameters that can be field sampled is provided in *Standard Methods for the Examination of Water and Wastewater* (APHA 1998) and USEPA (1996b), or the most current editions. An example of a field sampling program record sheet is provided in Appendix 6.

It is important to record the position of each sampling site so that it can be re-used in subsequent studies. It is also essential to make careful and thorough descriptions of the means of access and of the sites themselves and of the exact spots from which samples were taken. Key on-shore reference points should be identified, or the site should be located by the global positioning system (GPS), a satellite navigation system that enables points on the Earth's surface to be identified relatively accurately. Measurements by GPS are now becoming reasonably precise (to within 20 m). If exact positioning of sample sites is necessary, then basic GPS will be inadequate and a more accurate method (differential GPS) will be required. With high quality receivers and differential GPS the accuracy can be to within 1 m of the position or location. However, it is important to use a single coordinate system and to record which coordinate system is used, especially the datum and projection. A site identified by a latitude and longitude based on one datum can be up to 200 m away from a site identified by the same latitude and longitude numbers based on a different datum.

At each visit, the condition of the water body and the weather conditions must also be noted because these factors may influence the variables being measured. For example, changes in the wind speed and cloud cover may affect the temperature and subsequently the dissolved oxygen within the water column. Other field observations might include descriptions of odour, colour and floating material, and riverine vegetation or other conditions relevant to water quality. Video or photographic records are highly desirable for future reference.

4.3. Sampling of Waters and Sediments

4.3.1. Equipment and Methods

What are the most appropriate ways in which to actually collect samples or data from each sampling station and water body? Methods include:

- collection of a sample by hand,
- collection by automatic sampler,
- samplers that collect and integrate samples over a given time,
- real-time measurement by automatic means,
- measurements in the field by hand,
- remote sensing,
- field observation.

The choice of sampling method depends on the parameter to be measured and the nature of the information required. Differing sampling methods can provide differing information and have differing advantages and drawbacks. For example, grab samples could be easier to preserve, or less liable to contamination, or of a better size than samples integrated over time (or flow) by automatic devices. All the methods or equipment used must meet the relevant Australian and/or ISO Standard (e.g. AS/NZS 1998 a–e). This recommendation applies to all methods described in this document.

Selection of a sampling method should be guided by:

- the objectives of the monitoring program,
- the local conditions (i.e. the need to obtain representative samples),
- the safety of operation (the overriding principle should be the safety of the sampling staff),

- the acceptability of the method,
- commonsense.

Continuous sampling methods and equipment are being developed for waters and are able to operate reliably in some areas (Hart et al. 1993). They provide information on significant short-term variations in water quality parameters that are usually missed by discrete samples. Continuous sampling should be more widely used in the future as methods, equipment and data handling become more reliable. The procedures in this Monitoring Guidelines document nearly all involve individual sample collection, so they will need to be regularly assessed and updated or changed as technology improves.

Time-integrated sampling reduces analysis costs and enables mean values to be calculated simply. However, integrated sampling is not recommended where the objective is to assess variations in water quality.

Samples can be taken at the water surface, or at specific depths in the water column, or integrated over depths. For particular analytes (e.g. trace metals), the equipment must have a specific composition and be cleaned in certain ways to avoid sample contamination. A standard text for the general procedures and principles of collecting water samples is the *Standard Methods for the Examination of Water and Wastewater* (Methods 1060A and 1060B) (APHA 1998 or most current edition), mentioned above. Additional guidance on sampling lakes, rivers, streams, marine waters, groundwaters and sediments is found in the range of Australian and New Zealand Standards (AS/NZS 1998a–e, 1999).

The sampling operation includes the preparation and labelling of containers, appropriate selection of sampling sites, collection of samples, good housekeeping and field record books, photographic and video records, the use of boats and cars, and the recording of parameters such as depth and light intensity.

Green (1979), in his ten principles of sampling, says:

> verify that the sampling device is sampling the population you think it is sampling with equal or adequate efficiency over the entire range.

For this, the monitoring team must specify the population that is to be sampled and its likely spatial and temporal variability. In Australian rivers, discharge can change by two orders of magnitude, and the effectiveness of sampling devices may vary over this velocity range. Device-related sampling errors cannot be removed or accounted for by statistical methods or by replication, and in many cases they will be undetectable unless specific tests have been made.

The sampling device should not significantly disturb the environment being sampled or alter the samples taken, because if it does the samples will not reflect what ‘was’ or ‘is’. The problems in sediment sampling illustrate these difficulties. Blomqvist (1991) reviewed the problems of using several types of grab samplers and coring devices to obtain sediment samples. Grab samplers often do not enter sediments perpendicularly, and the sediment layers mix when they close. Most grab samplers have jaws that 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 must be designed to ensure that easily resuspended surface materials are not washed away. If rotation of cores occurs, shear stress may mix the sediment and cause core shortening.

Some consideration must also be given to the environment traversed by the sampling device, so that no sampling errors are caused 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 or it will pick up hydrocarbons from the water surface microlayer. On the other hand, when shallow water is being sampled, care should be taken not to stir up bottom sediment.

Sampling devices should be tested under controlled conditions to check that they quantitatively collect the sample of interest. In lieu of this, some studies reported in the literature compare the efficiency of sampling devices and document the limitations of various alternatives, e.g. water samplers (Harris and Keffer 1974), sediment samplers (Blomqvist 1991; Schneider and Wyllie 1991), biota samplers (Devries and Stein 1991). 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 sample site.

The sampling of waters for trace and ultratrace contaminants is increasingly a requirement for monitoring studies, especially for conformity with a guideline. To avoid sample contamination, much greater care is needed than for general water quality parameters. Non-contaminating equipment is essential, and it should be cleaned with acids for sampling metals, or cleaned with detergents and solvents for sampling organic compounds. Ahlers et al. (1990) elegantly describe the type of rigour required for preparing containers and for sampling methods. For trace metal surveys, avoid samplers with components that may contribute trace metals (Batley 1989). Use Perspex poles with all-plastic fittings to hold Teflon or polyethylene bottles for sampling shallow surface waters. Avoid depth samplers with rubber closures. For nutrient sampling, take care that samplers are free from residual nitric acid or phosphate-containing detergents that may have been used in their pre-cleaning. With samples for trace contaminants, the possibility of sample contamination is high; experienced staff may be required for such sampling. This is also true for samples of filtered nutrients where filtering is to be done in the field.

Many of the contaminants to be measured, particularly in relatively pristine marine or alpine waters, will be present at extremely low concentrations, which may influence:

- the volume of sample required (and hence the type of sampling device that may be suitable);
- the precautions required to avoid contamination (this could include the use of a suitable vessel such as a dinghy which can work away from the mother vessel, and the selection of sampling devices constructed out of non-contaminating materials); and
- the suitability of analytical methods.

Local conditions will further dictate the method and equipment used. To sample from a bridge a bucket can be used, while to sample from a river bank a telescopic pole would be more useful. For sampling in estuarine waters, the experimental design needs to take into account the complex and highly variable nature of the water body. In estuaries, waters intermix that have very different chemical composition and physical and chemical properties, producing great variation, vertically, horizontally and temporally (with tidal stage). As a result, large numbers of samples or stratified sampling may be required, which in turn is likely to have a bearing on the selection of the sampling method.

Details of specific sampling techniques for estuarine and marine waters are described by Grasshoff et al. (1999) and Crompton (1989), and can be found in the Australian and New Zealand Standard (AS/NZS 1998b).

4.3.2. Sampling of Surface Waters

Equipment for sampling surface waters (as opposed to groundwater) falls into five basic categories:

- bottle samplers for shallow waters,
- pumping systems for surface to medium (10 m) depths,
- depth samplers (50 m to >100 m depending on design),
- automatic samplers,
- integrating samplers.

4.3.2.1. Bottle Sampling of Shallow Waters

Many water bodies are shallow and well mixed, and surface (0–1 m) water sampling is all that is required. For this purpose, immersion of a sample bottle by hand to just below the surface (typically

0.25–0.5 m depth), is satisfactory, provided the hand wears a plastic disposable glove, and any contribution from surface films is avoided, and the sampler is downstream of where the sample is to be collected. This may be done from the shore, or by standing in shallow water, or from a boat. Whatever the vessel used it is important to face it into the ongoing current and to take water samples from the front of the vessel (Apte et al. 1998). This procedure minimises contamination from the boat itself. To maintain an adequate distance between the sampling point and the sampling vessel, the sample bottle can be held in acrylic jaws at the end of a 1–2 m long polycarbonate pole, 2 cm in diameter. This technique is also applicable for sampling from fixed structures. The alternative is a bottle or bucket fixed to a plastic rope. Both require cleaning before use and must be kept clean between uses by being housed in a plastic bag that is then placed in a clean plastic sealable container.

4.3.2.2. Pumping Systems

Pumping systems are effective for sampling, although they are not desirable for work on ultra-trace contaminants because the tubing gives the apparatus a large surface-area-to-volume ratio that increases the chance of adsorption of analyte. They are suitable for µg/L metal concentrations, and for all general water quality parameters. They typically involve a vacuum pump or, for shallower depths, a peristaltic pump. Water is sucked to the surface via appropriate lengths of pre-cleaned tubing of polyethylene or silicone or PTFE or PVC (approximately 1 cm diameter), and into a large acid-washed Pyrex glass Ehrlenmeyer flask (for a vacuum pump) or directly into a plastic sample bottle (for a peristaltic pump). The tubing is conditioned by pumping a large volume of water to waste prior to sampling. This procedure has been widely used for mercury sampling (USEPA 1996c) during which on-line filtration is also applied.

4.3.2.3. Depth Samplers

For depth sampling a range of purpose-built samplers are available (Batley 1989). Their basic operation involves a bottle, which can be opened at both ends via a wire or plastic line, which is deployed to the required depth. A weight is then sent down the line to trigger closure of the bottle and collection of the water. The bottle closures should be considered when determining the suitability of depth samplers; e.g. rubber closures are not recommended for low metal concentrations. In samplers for trace analysis, it is important to blank test the samplers by filling with them with clean water for the same length of time as the sampling event and then analysing the water for the analytes of interest. Contamination often increases with the age of the sampling device. The blank tests should therefore be carried out at regular intervals. The Mercos sampler, which uses a group of Teflon bottles, is ideal for obtaining water column depth profiles of trace metals including mercury at depths below 100 m.

4.3.2.4. Automatic Samplers

For unattended water sampling, an automatic sampler can be pre-programmed to collect samples continuously or on a flow-related or time-related basis. Such an arrangement is ideal for collecting stormwater runoff for example, and collection can be triggered by the commencement of water flow. A number of commercially available devices perform this function. They basically consist of a pump system, a controller and an array of sample bottles within a housing. Most instruments have purpose-made glass or polyethylene sample bottles that are designed so that a fixed number fit around the circumference of the housing. Glass is preferable for organic compounds and general water quality parameters, and polyethylene is better for metals. The bottles and all surfaces are acid washed before use. To check that the bottles are suitable for trace metal applications, clean water blanks are analysed after they have been in the bottles for a suitable test period.

Very large numbers of samples can be acquired using continuous samplers (AS/NZS 1998 b–d). The samples thus acquired can be bracketed, so that only a subset reflecting the conditions of particular interest need be analysed. It should always be borne in mind that the integrity of samples collected by automatic devices may be compromised by delayed preservation. This would normally require that samples be collected and processed as soon as possible after the relevant event. Refrigerated samplers

are also available, however, to assist in sample preservation. Automatic samplers may not be appropriate for sampling bacteria, pH or other variables that are likely to change significantly between the time of collection by the automatic sampler and retrieval from the field for analysis.

4.3.2.5. Integrating Samplers

For some sampling programs, single samplings are poorly representative of a site because water quality can vary considerably with time. In these cases samplers that integrate water samples over a fixed time period or volume are preferred. The samplers usually contain collector adsorbants that can quantitatively remove organic contaminants or metals, and are usually restricted to these parameters (Hart and Davies 1977; McLaren et al. 1985; Zhang and Davison 1995). The automatic samplers described above can perform as integrating samplers, although they are designed to collect a number of discrete samples.

Membrane-based passive samplers are a promising tool for the time-integrated monitoring of hydrophobic contaminants in aquatic ecosystems. In these devices, the uptake of chemicals is based on the partitioning of a compound between water and a lipophilic solvent enclosed in a semi-permeable polymeric membrane. Thus the passive samplers can be used as indicators of bioavailability of these contaminants. Laboratory analysis of the solvent is generally both faster and less expensive than water analyses for these compounds.

Several designs of membrane samplers have been proposed, including polyethylene bags filled with isooctane (Peterson et al. 1995). The main advantage is that they give a time-integrated measure of toxicant exposure at sites that can be related to any observed ecological changes. For example, endosulfan runoff from cotton fields measured by in situ passive samplers was found to be a good predictor of changes in density of benthic organisms (Leonard et al. 1999). Passive samplers have been adopted as a quantitative sampling method for pesticides in New South Wales to detect chemicals that standard sampling methods may have missed (Muschal 1998).

4.3.3. Sampling Groundwaters

Groundwater occurs in aquifers at various depths below the ground. Recharge may be by direct infiltration of rainfall, by seepage from rivers or other bodies of surface water, or by transfer from one aquifer to another. The area of recharge may be at the sampling site or many hundreds of kilometres away. The water may have been resident in the aquifer for a few days or millions of years.

The quality of groundwater can vary from almost pure water to extremely concentrated brines. Its quality depends on the geology of the aquifer and can be subject to contamination from substances that come into contact with the ground. Fertilisers, pesticides, petroleum products, landfills, mining, household and farm and industrial wastes all contaminate groundwaters to varying degrees, often much more than surface waters.

Monitoring of the quality of groundwaters involves techniques different from those used for surface water quality investigations because groundwater, by its very nature, cannot be sampled without some disturbance from the construction of a bore or other access hole and the effects of sampling devices and procedures. These may also cause chemical and biological contamination unless stringent precautions are taken. Hence sampling staff must make extreme effort to ensure that the samples are representative of the water in the aquifer. Groundwater sampling should generally be carried out by experienced field staff or in close consultation with experts, to ensure sample integrity.

To retrieve a representative sample, these principles should be considered (see also Table 4.2 and QDME 1995):

- the sampling equipment should not change the water quality in any way; particular effort should be made to avoid cross-contamination between bores and sampling equipment;
- sufficient water should be removed to ensure the sample is newly derived from the aquifer itself rather than from water that has sat in the bore; and

- the methods of collection and storage in bottles and transportation to the laboratory should be suitable for the type of analysis required.

Guidance is provided in the Australian/New Zealand Standard for groundwater sampling (AS/NZS 1998e).

Groundwater sampling may produce a large volume of purged water. If necessary, this should be stored on site in drums for proper disposal.

Table 4.2. Groundwater samplers (from QDME 1995)

Type of sampler	Remarks
Displacement pumps	These types of pumps provide a gentle pumping action which is suitable for purging and producing a high quality sample for all reasonable purposes.
(i) Positive or gas bladder displacement	If the pump is made from sufficiently high quality materials, the types of media used to work the pump are not important.
(ii) Mechanical displacement	These pumps are very suitable for well purging, though they are often slow. Specific design is very important in performance.
Submersible pumps	Several varieties of this type of pump are available, from very cheap to quite expensive. They rely on a motor below water level powering a pump to push water to the surface via a delivery line.
	These pumps are efficient in purging bores and can produce an acceptable sample for most purposes, though minor gases may be lost.
Suction pumps (centrifugal)	An excellent method for purging wells, but limited to about 6 m depth. The 'suction' type action has little effect on the water remaining in the well, though the samples pumped may lose some gases or organic compounds. Following purging with this type of pump, high quality samples can be taken with balers, with care.
Down-hole grab samplers	Of little use in sub-artesian bores. Can obtain high quality samples from different depths in flowing bores. Quality control must be exercised in the transfer of samples.
Balers	Difficult or impossible to use for purging bores. Require extreme care to prevent contamination, because samples have to be pulled to surface. Rope needs to be sterilised, cleaned or replaced frequently.
	Good quality samples can sometimes be obtained if bores are purged by other approved means. Quality control must be exercised in transfer of samples.
Air-lifting	This is generally considered to be a poor method of obtaining high quality samples. The air-lifting method strips gases and organic compounds from the water, changes pH, and may cause minor chemical changes. However, it is efficient in purging bores, and has little if any effect on major ion chemistry. Generally, results of analyses of air-lifted samples are considered acceptable for monitoring major ions.
	Samples should not be submitted for higher quality types of analyses.
	This method is not considered suitable for well purging for high quality samples even if sampled with a baler. Air-lifting has an effect on the whole water column and the whole volume would have to be removed before a high quality sample could be obtained.

4.3.4. Sampling of Precipitation

This Monitoring Guidelines document does not cover collection of deposited rain, snow and airborne particulates in detail, although the analytical methods described in Chapter 5 are applicable to these types of samples. The user is directed to the WMO *Manual on Water Quality Monitoring — Planning and Implementation of Sampling and Field Testing* (1988). The Commonwealth Bureau of Meteorology and the relevant state and territory agencies that undertake sampling for these measurement parameters should be consulted for guidance and additional information.

4.3.5. Sediment Sampling

Sediments often are surveyed to determine the composition and concentration of contaminants in them, as well as the numbers of organisms located at various depths. There are two broad-based sediment classifications: *suspended sediments* and *bottom sediments*. In water quality terms, suspended sediments are generally dealt with as part of the water column, although specialised sampling techniques are required for obtaining representative samples (USEPA 1991b; AS/NZS 1999). The benthic organisms in bottom sediments are investigated as measures of aquatic health, pollution or contamination, and as part of the ecology of aquatic systems. In this section we are only concerned with sampling the sediment, not the benthic organisms within the sediment. The monitoring team must decide what to include in the sample before beginning sampling. For example, when the objective is to sample sediments from within a seagrass bed, it is normal practice to remove the rhizomes of the seagrasses and to sieve sediments to remove small molluscs.

It is recommended that sampling for sediments should follow internationally accepted protocols and procedures. The choice of sampling method will be dictated largely by the nature of the investigations being undertaken. However, it may be necessary to consult with the relevant state and territory agencies about any internal guidelines that they may have for sampling sediments. Table 4.3 lists a number of sediment sampling methods that can be used both for biological and for non-biological parameters. Sampling for determination of sediment transport is not covered in the Monitoring Guidelines.

Table 4.3. Methods for sampling sediments

Method	Reference
Method 9060A Sample collection	APHA (1998)
Method 10500B Soft-bottom dredge	APHA (1998)
Method 10500B Hard-bottom dredge	APHA (1998)
Method 10500B Rocky-bottom samplers	APHA (1998)
Dredging method	APHA (1998)
Grab method	EPAV (1992)
Corer method	EPAV (1992)
Integrating samplers	EPAV (1992)
Grab samplers	WMO (1988)
Core samplers	WMO (1988)
Protocols for sampling	Environment Canada (1994)

For most applications, sediment coring is recommended (Mudroch and Azcue 1995). With this technique, samples can be taken to a measurable depth and then subsampled to provide depth profile information. Corers generally vary in diameter from 2.5 cm to 5 cm, and range from long PVC pipe (2–3 m), that can be immersed in shallow waters from a boat, to shorter Perspex, polycarbonate or other tubes that can be immersed by hand by divers. The tubes have a bevelled leading edge to ease their movement through the sediment. In shallow waters, cores can be extruded by gas pressure and subdivided, but this is not recommended where sediment oxidation is an issue. If divers are not available, vibrocorers are essential for use from vessels in waters deeper than about 3 m. Vibrocorers usually contain plastic liners that protect the sample from contamination. Some corers enable in situ freezing of the sample.

A grab sampler or dredge is a useful alternative for obtaining large volumes of surface sediments. A range of types of dredge sampler is available and most of them are suitable for sampling in shallow water depths (<20 m). However, care should be taken in deeper waters to ensure that fines are not lost during the passage of the sampler to the surface (APHA 1998), because it is these particles that are most enriched in trace contaminants (Mudroch and Azcue 1995).

When the chemical forms of contaminants and their associations with sediment phases are to be determined, it is necessary to ensure that the redox state of the sediments (oxic or anoxic) is not altered, because oxygenation (or reduction) will cause irreversible changes. Sediments become oxygenated on contact with air, so the sediment samples need to be capped immediately at sampling and stored in a nitrogen glove box. Oxidation can be minimised if samples are frozen at –20°C.

4.3.6. Sample Containers

The monitoring team must decide on appropriate sample containers (type and volume) and how to clean them before use. The sample container can affect the composition of the sample by adsorbing some of its constituents (Batley 1989); for example, glass containers tend to adsorb phosphate. On the other hand, as mentioned above, containers can also be a source of contamination unless they are carefully prepared.

Metals can be present at trace concentrations both on glass and on plastic surfaces, while organic compounds are more likely to be found on plastic containers. Bacteria on container walls may use nutrients from solution (Maher and Woo 1998). The caps of containers often contain inserts of cardboard, cork or rubber that should be removed because they can cause contamination. A range of plastic materials have been used and their propensity for sample contamination has been thoroughly reported (Hunt and Wilson 1986; Hall 1998; Reimann et al. 1999). For metals, the preferred sample containers are fluorocarbon polymers, PTFE (Teflon) or FEP, as well as high-density polyethylene. Bottles made of FEP are usually only used for mercury analysis because they are so costly. High quality bottles are recommended, e.g. Nalgene, because these have good closures that prevent sample leakage. For samples to be analysed for selenium, bottles made of polycarbonate and some types of polyethylene are not suitable. For nutrients, polyethylene (low density or high density) sample bottles are the most favoured type. Glass is not favoured because there can be high concentrations of trace metals in the glass and there is potential for adsorption losses.

Some degree of cleaning is usually applied to bottles before they are used. This often involves soaking in acid, but the rigour of the procedure varies from laboratory to laboratory. Some authors have advocated direct use of certain bottle types without any cleaning (Reimann et al. 1999). Such sweeping statements are ill-advised because the quality of sample bottles often changes between batches. In the authors' laboratories, the bare minimum cleaning procedure involves soaking sample bottles in 10% nitric acid for at least 24 hours and then rinsing them with copious quantities of deionised water. Such precautions are worthwhile on most occasions, given the cost of sampling (especially if helicopters or boats are involved in the sampling programs). In our laboratories, acid washing is carried out in a dedicated dust-free room, and the acid baths are stored in a bunded and vented area similar to a very large domestic shower recess. The emptied bottles are double-bagged using two zip-lock polyethylene bags. For waters for zinc analysis, Ahlers et al. (1990) advocated that Nalgene bottles be soaked in hot 50% nitric acid for two days, rinsed with high purity water, then leached in 1% nitric acid for two weeks. The value of such extreme care was clearly demonstrated in the reliability of the resultant analytical data. In any case, sampling staff should check with their analytical laboratory to ensure bottles have been appropriately prepared before use.

4.3.7. Sampling Protocols

When sampling waters containing trace metals, nutrients or organic compounds, a single person wearing plastic disposable gloves and taking appropriate care can carry out the operation without sample contamination if he or she is alert to potential sources of contamination. Lavish use of polyethylene sheeting to wrap equipment and to cover work areas on boats, river banks, etc., is part of the good practice that follows automatically from this alertness. Dust, powder, skin and hair are obvious external sources of metals, and rigorous care is required to minimise their effects. Detailed protocols for ultra-trace sampling have been described in the literature (Ahlers et al. 1990; Nriagu et al. 1993; Nolting and de Jong 1994; Apte et al. 1998); see also section 4.6.2 below.

The recommended sampling protocol for ultra-trace analysis uses the 'dirty hands–clean hands' approach. This involves two people both wearing (powder-free) polyethylene gloves. The double-bagged bottles are sequentially unwrapped. The first 'dirty' assistant removes the outer bag and hands the bottle to the 'clean' assistant, who removes the inner bag. The 'clean' assistant then immerses the bottles by hand or from the end of a pole (as discussed above), or fills them with the sample collected using the depth sampler. Bottles are usually rinsed with sample material first — filled, capped, shaken and emptied —before being refilled with the sample to be analysed.

Other precautions also help avoid contamination: reagents for use in the field are stored in decontaminated containers; sample and reagent containers are transported in separate sealed plastic bags; all field equipment is pre-cleaned to the same standard as the containers; containers are uncapped or removed from their transport bags for minimum amounts of time; containers that were filled with water as part of the preparation protocol are emptied well away from and downstream of the sampling location before being rinsed with sample and refilled.

4.4. Sampling of Aquatic Organisms

For aquatic organisms, selection of a sampling method should be again be guided by:

- the objectives of the monitoring program;
- the local conditions (i.e. the need to obtain representative samples);
- safety of operation (the overriding principle should be the safety of the sampling staff);
- acceptability of the method; and
- commonsense.

The organisms typically sampled comprise plankton, bacteria, periphyton, protozoa, algae, fungi, macrophytes, macroinvertebrates, benthic macroinvertebrates and algae, bivalves and fish. Methods for their sampling are described in APHA (1998 or most current edition) and Hellawell (1986) and summarised in Table 4.4.

The monitoring team will have decided which organisms to collect at the study design stage (section 3.5.2 in Chapter 3) and now must choose the appropriate equipment and procedures to use. Several devices may be needed to ensure quantitative sampling of all the required organisms. There may need to be a compromise between these and more rapid methods. Devries and Stein (1991), in their comparison of the efficiency of three sampling devices (tube sampler, vertical tow net and Schindler–Patalas trap) for collecting zooplankton, found there was no best method. Zooplankton consist of a mixture of copepods, cladocerans and rotifers. Generally copepods and cladocerans were best collected using the tube sampler, while rotifers were best collected using the Schindler–Patalas trap. Some species were best collected using the vertical tow net. It is important to decide which biota is to be collected before sampling begins. Aquatic biologists should be asked for specific guidance about the best way of sampling this biota. As indicated above, there is no universal method for collecting biota and if a range of biota is collected, several sampling procedures can be needed.

Table 4.4. Methods for sampling aquatic organisms

Organism sampled	Method	Reference*
Plankton	Grab sample/plankton (cone) nets, hose-pipe sampler, Patalas–Schindler plankton trap	APHA (1998), Hellawell (1986)
Bacteria	Grab sample	APHA (1998), Ward and Johnson (1996)
Periphyton	Artificial and natural substrates (Paddlepop sticks, modified brushes)	APHA (1998)
Protozoa	Grab samples	APHA (1998)
Algae	Grab samples/nets, hose-pipe samplers	APHA (1998), Falconer (1994), Hotzel and Croome (1998)
Fungi	Grab samples	APHA (1998)
Macrophytes	Various	APHA (1998)
Benthic macroinvertebrates and algae	Bottom grabs/samplers/diver-held cores/nets	APHA (1998), Growns et al. (1999)
Macroinvertebrates	Sweep nets, hand search, quadrats, long-handled pole with net (deep waters)	APHA (1998)
	For shrimps, crayfish: gee-traps, net traps	
Bivalves	Cage, basket sampler, by hand	APHA (1998)
Fish	Nets, traps, electro-fishing	APHA (1998), Harris and Gehrke (1997)

*for APHA (1998) use this or the most current edition of this text

4.5. Sample Preservation and Storage

In most cases, samples are collected for later chemical or biological analysis. In all cases, clear and distinctive sample labelling is important. After collection, it is important to maintain the integrity of each sample and to ensure that it does not become contaminated or change between collection and analysis. It is usually necessary to preserve the samples to retard biological, chemical and physical changes. Protocols must specify both the appropriate sample container and the preservation technique. Preservation choices will vary depending on the parameter to be measured. Some possible changes, and suitable preservation strategies or storage procedures are given in Table 4.5.

Matters for consideration to ensure successful 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 available will depend on the variable to be measured. Comprehensive information on the selection of containers and preservation of water samples for chemical and microbiological analysis can be obtained by consulting the Australian Standards (AS 1987a,b). Complete and unequivocal preservation of samples is a practical impossibility. At best, preservation techniques only retard the chemical and biological changes that inevitably take place after sample collection.

Normally, to prevent chemical and biological changes, water samples are cooled to 4°C, or frozen, or filtered, or given a chemical additive. Freezing (–10°C) reduces, but does not eliminate, biological activity in samples. All biological activity is only effectively eliminated at –40°C. Chemicals such as chloroform and mercury (II) acetate have also been used to prevent biological activity. Acid is often added to prevent adsorption of metals from water samples to containers and precipitation of insoluble salts.

Table 4.5. Preservation and storage strategies for physical, chemical and biological samples

Change	Preservation techniques
Physical	
Adsorption/absorption	Inorganic: reduce pH on storage
Volatilisation	No head space
Diffusion	Choose correct container type and cap liners
Chemical	
Photochemical action	Use dark containers
Precipitation	Lower pH, avoid use of chemicals which cause precipitation (e.g. sulfates)
Speciation	Refrigerate at 4°C. Add fixing agent
Biological	
Microbiological action	Reduce pH, filter, add bactericide, e.g. for sulfide add zinc acetate; if chlorine present add thiosulfate, leave small airspace to preserve viability, avoid light, refrigerate (4°C)
Cell degradation	Freeze, add fixing agent, e.g. formaldehyde, ethanol

Chemical preservatives should be avoided, if possible, because they may contaminate samples or interfere in chemical or biological analysis. For example, mercury can interfere in the colorimetric determination of phosphate. If preservatives are used they should also be taken into account in the analysis of blanks.

Even if a sample is frozen or a preservative is added, samples can be stored only for a finite time. In some cases this period may be years, e.g. phosphorus in seawater, but in other cases it may be much shorter, e.g. six hours for *Escherichia coli* samples. The preservation time needs to be determined before samples are collected, and protocols must be designed to ensure that samples are analysed before a significant change in composition occurs.

4.6. Quality Assurance and Quality Control in Sampling

A quality assurance and quality control (QA/QC) program for field sampling is intended to control sampling errors at levels acceptable to the data user. Thus it includes procedures designed to prevent, detect and correct problems in the sampling process and to characterise errors statistically, through quality control samples. Major errors to be avoided are faulty operation of the sampling device, changes in the sample before measurement (contamination, chemical or biological changes), and incorrect sample labelling.

Field staff should be competent in sampling and making field measurements even though they may also have qualities, such as vehicle handling or bush skills, unrelated to the assurance of sample integrity. Before sampling staff are permitted to do reportable work, they should demonstrate competence in field procedures. As a minimum this would include being able to adhere to protocols, being able to avoid contaminating samples, and being able to calibrate field instruments and make field observations.

All equipment and field instruments should be kept clean and in good working order, and calibrations and preventative maintenance should be recorded carefully. All repairs to equipment and instruments should be noted, as well as any incidents that could affect the reliability of the equipment. When automatic sampling devices are used, their timing mechanisms must be calibrated to ensure that the samples are acquired at the specified intervals. This is especially important where hydrological or other conditions result in significant short-term concentration variations.

4.6.1. Tracking Samples and Field Data

During sampling or field measurements, it is important to fill in a field data sheet or similar record that describes the samples taken, their labels and other details about them (e.g. see Appendix A6.1). All field data and instrument calibration data are recorded on this sheet. All field records must be completed before leaving a sampling station. Any observations or information on the conditions at the time of sampling that may assist in interpretation of the data should be noted on a field-record sheet or in a field notebook. This information may explain unusual data which otherwise might be attributed to problems in sampling or analysis.

Table 4.6. Chain of custody documentation

Process step	Quality Assurance Procedure
Field sampling	Field register of sample number, site, type/technique, time, date, technician, field data sheet
Sample storage and transport	Field register of transport container number and sample numbers, time, date
Laboratory receipt of samples	Laboratory register of transport container number and sample numbers, time, date
Laboratory storage of samples	Laboratory register of storage location, type, temperature, time, date
Sample preparation	Analysis register of sample (laboratory) number, pre-treatment, date, technician
Sample analysis	Analysis register of instrument, calibration, technician, standard method, date, result

If samples are to be the basis for legal proceedings at some time in the future, the following questions are likely to be asked:

- exactly where was the sample taken?
- was the person who took the sample competent to do so?
- how was the sample labelled to ensure no possibility of mix up or substitution?
- was there any possibility of contamination, e.g. of the container, of the sample during filling, or later?
- did the sample deteriorate after collection?

Chain of custody documentation (Table 4.6) ensures that these questions can be answered.

4.6.2. Documented Sampling Protocols

Sampling errors can be minimised by ensuring that correct procedures have been followed during the field sampling, transport and storage. Sampling protocols need to be written and adhered to: they must include detailed descriptions of the procedures for collecting, labelling, transporting and storing the samples and necessary ancillary field data. Protocols must be specific to each matrix and constituent, and specify the sample collection device, type of storage container and preservation procedures.

The protocol must also specify the types and numbers of quality control samples to be taken. Before this protocol can be written, the nature of errors, both systematic and random, and the level of accuracy desired must be assessed. Sources of error include reaction with sample or sample container, contamination (field, sampling device, containers), chemical and physical instability, and biological changes.

The exact locations of sampling sites and any sub-sites must be recorded in the sampling protocol. Field notes must accurately describe where samples were collected, to allow cross-checking with the

sampling locations specified in the sampling protocol. If transects are to be sampled the location range must be specified if this is within the precision of the positioning instrument. Taking note of the time when samples are taken (standard or daylight-saving time) is an obvious but frequently overlooked requirement of rigorous sample definition (Appendix 6).

Protocols should specify how sampling staff are to be trained to use sampling equipment. Problems that may occur in the field should be anticipated: sample containers may be lost; sample volumes may be low; should foreign objects be included? on the basis of what criteria is foreign matter rejected? what if sites cannot be sampled?

One of the major challenges of sampling is to prevent contamination. Protocols must include the following basic precautions for avoiding contamination:

- field measurements should be made on separate sub-samples of water;
- new or reused sample containers must be appropriately cleaned (use of containers supplied by the analytical laboratory is recommended);
- only the sample bottles recommended for each parameter should be used;
- container lids should be checked for liners that may cause contamination or adsorb particular analytes;
- containers that have already been used for other purposes should be discarded;
- the insides of containers and lids should not come in contact with hands or objects;
- sample containers and filter units should be kept in a clean environment away from dust, dirt, fumes, etc.;
- preservatives should be tested for contamination;
- care should be taken to avoid cross-contaminating samples when adding preservatives;
- sample containers used for collecting samples for microbiological analyses must be sterilised;
- sampling staff should use plastic disposable gloves when handling sample containers at every stage during sampling (to avoid touching the sample, and the insides of caps or containers).

4.6.3. Sample Blanks and other QA/QC Practices

4.6.3.1. Blanks to Check on Field Procedures, Containers, Equipment and Transport

If it is possible that there could be contamination during the sampling process, blank samples should be devised to detect and measure the contaminant.

For field blanks, extra containers with suitable contents are taken to the site. There, the containers are opened and closed and the contents are handled just as if these were real samples during transfer and storage. For freshwaters, sample bottles filled with deionised water are used as field blanks. For marine work, water of the appropriate salinity is used. One blank per 10 samples is prepared, adding any preservative in the field. Field blanks mainly detect contamination from dust and other atmospheric fallout.

Filter blanks allow estimation of contamination by filtration in the field. They are prepared in the field by passing a sample of distilled water through a pre-cleaned filter and adding preservative to the water sample.

Container blanks determine the contamination from the container. Containers of each type to be used for sampling (about 1 in 10) are selected at random and filled with deionised water and preserved in the same manner as field samples. Analysis of these blanks detects contamination by the container washing process. This is sometimes measured as a rinse blank, and for this the last of several distilled water rinses of sampling equipment in the field is analysed.

Equipment blanks measure contamination introduced through contact with sampling equipment or sampler. They consist of the water or solvent that is used to rinse the sampling equipment between samples.

Trip blanks can be used to assess gross cross-contamination of samples during transport and storage. The simulated samples are similar to the samples to be collected, but in them the substance to be analysed (analyte) is at background or low concentrations.

Often it is not possible to achieve no contamination, but stable contamination levels are aimed for instead. When levels of contamination are outside the agreed acceptable limits, the contamination is likely to be coming from new sources.

Checks for QA/QC are partly reactive. If changes in samples are detected by using standard additions or blanks, a specified procedure is devised to determine and rectify the problem. The water is re-sampled if possible.

4.6.3.2. Duplicate Samples

Besides blanks, other sample QA measures include the use of multiple samples. Duplicate samples reveal the magnitudes of errors (contamination, random and systematic) occurring between sampling and sample analysis. They are obtained by dividing a sample into two or more sub-samples. On the other hand, replicate samples are two or more samples collected simultaneously to establish the reproducibility of sampling. Ideally three samples are required to enable testing of inter- and intra-laboratory accuracy and precision.

4.6.3.3. Sample Spiking

Another alternative is to 'spike' sub-samples in the field to detect change: a known amount of the analyte of interest is added to the sub-sample and subsequently measured. Samples for QA/QC should be labelled in such a way that they are not distinguishable from other samples in the batch.

4.6.4. QA/QC in Biological Sampling

The main question to be addressed for biological sampling is whether it is quantitative and representative. Alternative sampling strategies need to be devised and tested to establish the suitability of any preferred sampling technique.

4.6.5. QA/QC in Data Storage and Access

Transfer of results from the field to a database should be automated where possible, and the printout of the entry should be checked against the field record sheet and the laboratory register. Entries can be validated by electronic screening against the expected range and against other analytes for the same site and sampling date, and against field measurements. There should be agreed procedures for handling and tracking updates and corrections to data. There should be provision for handling censored data (see section 6.2.1 in Chapter 6). There should be fields for all necessary identifiers, for traceability purposes, e.g. sample and laboratory numbers.

Quality assurance also relates to data security and backup. With respect to security, those personnel who have read- or write-access to the data must be specified. Data backup is always essential in case of system or file failures.

4.7. Occupational Health and Safety

4.7.1. Identification of Hazards

Hazards or risks involved with field sampling need to be identified and documented on a preliminary site visit. The major questions to be resolved are these:

- can staff reach the site in safety?

- can a sample be safely taken? Is the water fast flowing? Is a boat to be used? Is there safe boat access? Is the site prone to flash floods? Is the bank stable? Are tidal changes likely?
- will sampling staff be exposed to toxic or other hazardous substances?
- will sampling staff be exposed to any pathogens, e.g. Ross River virus, malaria, etc.?
- will any potentially dangerous fauna be encountered, e.g. spiders, ticks, snakes, leeches, crocodiles, sharks, pigs, etc.?
- are weather conditions likely to endanger personal safety? In alpine areas especially, weather patterns are extremely variable.

Personnel who are to conduct sampling should be physically and mentally able to carry out field work. For example, if sampling staff fall into a water body, they must be physically fit enough to get out without assistance (although staff should never work alone in the field). Sampling staff working near water must be able to swim. They must also be able to climb up river banks. In proper professional practice, risks must be reduced as much as possible, and staff must not be required to operate in conditions that are unsafe.

4.7.2. Education About Hazards

All staff must be appropriately trained as part of the formal risk minimisation strategy. Training will include:

- familiarisation with environmental hazards that may be encountered;
- familiarisation with sampling protocols (sampling procedures, chain of custody considerations, etc.);
- use of sampling equipment;
- qualifications to drive appropriate vehicles, e.g. off-road 4-wheel-drive vehicles, bikes, tractors or boats;
- familiarisation with safety procedures;
- qualifications in advanced first aid.

4.7.3. Risk Minimisation Plans

The following directives should reduce risks during sampling operations.

- Limit continuous driving. If sampling sites are at a considerable distance, do not drive there without a stop. Take breaks of at least 15 minutes every 3 hours, and sample for no more than 10 hours in one 24 hour period.
- Choose safe sites with safe access. Visit potential sites and check them after they have been tentatively selected from map surveys. They should have reasonable access, be free of dangerous animals or prickly or poisonous plants, have no steep, slippery or unstable banks, and not be prone to rapid flooding or tidewater rise without warning.
- Wear appropriate clothing. Obtain weather forecasts for an area to be sampled. Be prepared, for example, to wear raincoats if there is likelihood of rain, warm clothing if it is cold, hat and sunscreen at all times, and footwear with a good grip for wet rocks (do not go barefoot, risking injury from sticks or broken glass). Note that sunscreen can be a source of contamination and should be used with due care for this reason. Take extra clothes and a towel in case of someone falling in the water.
- Take appropriate safety gear and a first aid kit. Wear lifejackets when sampling near deep water with poor footing or from a boat. Plastic gloves are essential to anyone who has an open or bandaged wound when handling chemicals or contaminated water or even if the water quality at the site is unknown. Take a fully stocked first aid kit to the monitoring site; ideally, someone in the monitoring party should have first aid training.

- Maintain contact with help and never sample alone. Work with at least two others and stay in contact with someone who can raise the alarm; carry a mobile telephone if available or at least keep coins or phonecard to be able to make a telephone call. In remote areas, carry maps, compass, mirror and matches and inform a responsible person of intended movements. There must be written procedures describing how emergency services are to be accessed.
- Never go into deep water. Sampling in deep water requires the use of an appropriate boat with the necessary safety equipment (life jackets, flares, etc.); preferably, sample from a bridge or use a cableway if installed at the site.
- Avoid contact with contaminated water. Carry drinking water. Do not drink from the source being monitored. Always wear plastic gloves when water quality at the site is unknown and in particular when collecting samples in which the presence of algae, pathogenic organisms or toxins can be expected (blue–green algae can cause skin and eye irritations). Wash hands after monitoring and before eating; treat all bacterial cultures as pathogenic.

Professional practice also requires sampling staff to:

- obtain approval as required, such as permits to collect fauna and flora or take water samples;
- have access to sites; land-holders' permission may be required to enter private land;
- use appropriate etiquette. It is good practice to inform local authorities, park rangers, etc., even if formal permission is not required. Local people can give useful information that helps in the choice of safe sampling locations and warns of local hazards.

Individual sampling staff have a duty of care to other personnel. Considerations include these:

- if one person cannot carry out all aspects of field work then he or she must have colleagues to assist;
- there should be no discrimination;
- privacy of individuals should be respected.

There is a duty of care to avoid damaging the environment during sampling:

- do not litter;
- observe fire restriction requirements;
- do not wash in streams, lakes, estuaries, etc.;
- remove human wastes;
- do not feed native animals;
- minimise environmental damage by keeping to paths, tracks, etc.

Chapter Five
Laboratory Analysis

5.1. Introduction

The aim of laboratory analyses is to obtain accurate and precise data in a safe environment. A framework for designing an analysis program is given in Figure 5.1. A checklist of important considerations is given in Table 5.1. For a typical laboratory request form see Appendix A6.2.

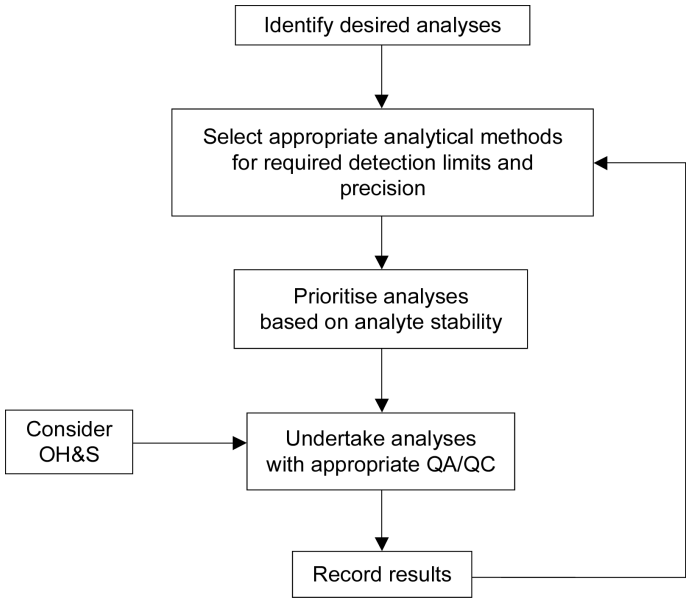


Figure 5.1. A framework for designing an analysis program

5.2. Analytes

The particular substances to be analysed (analytes) are the focus of the monitoring program. They may have been identified in generic terms during the study design, but now the individual compounds need to be decided on, and possible methods of determination need to be considered before planning the laboratory program. The analytes will also determine many of the decisions involved in laboratory analyses: for example, how to obtain good quality data (method and equipment), how to protect the health and safety of workers, and how much this stage of the monitoring program will cost.

Table 5.1. Checklist for undertaking laboratory analyses

1. Have the analytes been clearly stated?
2. Have appropriate analytical methods been identified?
   (a) Will analytical methods cover the range of concentrations expected?
   (b) Will analytical methods detect the minimum concentration of interest?
   (c) Will the analytical methods have sufficient accuracy and precision?
   (d) Will the substances be processed within the samples' storage life?
3. Does the laboratory have the appropriate equipment to undertake the analytical method chosen?
4. Are laboratory facilities (water supply, air supply, environment) suitable for the planned analyses?
5. Do the laboratory staff have the expertise, training and competence to undertake the planned analyses?
6. Has a laboratory data management system been established? Does the system
   (a) track samples and data (chain of custody)?
   (b) have written data entry protocols to ensure correct entry of data?
   (c) enable associated data to be retrieved, e.g. nutrient concentrations and flows to calculate nutrient loads?
   (d) have validation procedures to check accuracy of data?
   (e) have appropriate storage and retrieval facilities to prevent loss of data and enable retrieval (for at least three years) based on current and unexpected information needs?
   (f) Are procedures in place to ensure information has reached the user?
7. From the documentation, can this information be seen:
   (a) how results were obtained?
   (b) that samples had unique identification?
   (c) who the analyst was?
   (d) what test equipment was used?
   (e) the original observations and calculations?
   (f) how data transfers occurred?
   (g) how standards were prepared?
   (h) the certified calibration solutions used, their stability and storage?
8. Has a laboratory quality assurance plan been developed? Does the plan specifically cover:
   (a) operating principles?
   (b) training requirements for staff?
   (c) preventative maintenance of laboratory infrastructure and equipment?
   (d) the requirements of the laboratory data management system?
   (e) procedures for when and how corrective actions are to be taken?
   (f) allocation of responsibility to laboratory staff?
   (g) all documentation to maintain quality at specified levels?
   (h) quality control procedures to minimise analysis errors?
   (i) quality assessment procedures to determine quality of data?
9. Are all protocols for preparing and analysing samples written and validated?
10. Are standard methods being used? If variations of standard methods or non-standard procedures are used, procedures must be technically justified, with documentation of the effects of changes.
11. Have analytical methods' accuracy, bias and precision been established by:
    (a) analysis of standards?
    (b) independent methods?
    (c) recovery of known additions?
    (d) analysis of calibration check standards?
    (e) analysis of reagent blanks?
    (f) analysis of replicate samples?

Table 5.1. continued

12. Do operation procedures specify instrument optimisation and calibration methods, and do these cover:
    (a) preventative maintenance?
    (b) specific optimisation procedures?
    (c) resolution checks?
    (d) daily calibration procedures?
    (e) daily performance checks?
13. With respect to laboratory QA/QC:
    (a) have control charts for laboratory control standards, calibration check standards, reagent blanks and replicate analyses been established?
    (b) does the laboratory participate in proficiency testing programs?
    (c) does the laboratory conduct unscheduled performance audits in which deviations from standard operating procedures (protocols) are identified and corrective action taken?
14. Have all reasonable practical steps been taken to protect the health and safety of laboratory staff?
    (a) have hazards been identified?
    (b) have laboratory staff been educated about hazards?
    (c) have risk minimisation plans been prepared?
    (d) have staff been trained to ensure safe work practices?
    (e) are staff appropriately supervised?
    (f) are staff insured?

5.3. Choice of Analytical Methods

The selection of an analytical method for waters, sediments or biota will largely depend on the information and management needs of those undertaking the investigation, and on the analytes themselves. However, limitations such as the financial resources available, laboratory resources, speed of analyses required, matrix type and contamination potential, are also important factors.

The choice of an appropriate analytical method is based on four considerations:

- the range of concentrations of the analyte that need to be determined. Detection limits are method specific and the lowest concentration of interest will need to be specified.
- the accuracy and precision required. All results are only estimates of the true value and the greater the accuracy and precision required the greater the analytical complexity and cost.
- the maximum period between sampling and analysis. On-the-spot field analysis may be required, depending on the use to be made of the data.
- Where several methods can achieve the above requirements, the ultimate choice may be dictated by familiarity with the method and/or the availability of necessary analytical instrumentation.

Appropriate procedures for both chemical and biological analyses can be found by reference to accepted published procedures such as *Standard Methods for the Examination of Water and Wastewater* (APHA 1998) or USEPA sampling and analysis methods (Keith 1991), or the most recent editions of these. Methods that cover water, sediment and biological analyses are available on the USEPA web site, http://www.epa.gov. Detailed reviews or descriptions of methods are available in standard texts; however, a good general text with methods encompassing all the parameters covered in this section is Rayment and Higginson (1992).

Methods for marine water analysis are found in Grasshoff et al. (1999) and Parsons et al. (1985). A summary of analytical methods and associated references is provided in Table 5.2. Methods for toxicity testing are not included, but have been summarised in section 3.5 in Chapter 3. Standard methods are updated regularly; however with the pace of research, many good non-standard methods are available that have not yet been included. Their use is acceptable, provided that justification for

their choice is given and that their performance can be demonstrated through the analysis of standard reference materials or other quality control procedures.

Aspects or details of both standard and non-standard procedures may require evaluation or modification for use in Australian or New Zealand conditions. Some of these modifications are well documented, but others either are not recorded or are recorded in commercial methods that are not available to the authors of the Monitoring Guidelines.

Table 5.2. Summary of analytical methods for physical and chemical parameters

Analyte	Methods	References*
Physical		
Clarity	Secchi Disk	APHA (1998)
Temperature	Thermometer, electronic data logger, thermistor	
Flow rates	Acoustic Doppler current profiler, assorted methods	USEPA (1982), RD Instruments (1989)
Depth	Acoustic Doppler current profiler, depth sounder	RD Instruments (1989), EPA (1992)
Colour	Colorimetry	APHA (1998), USEPA (1983)
Suspended solids	Gravimetry	APHA (1998), USEPA (1983), AS (1990)
Turbidity	Nephelometry, light scattering	APHA (1998), USEPA (1983)
Conductivity	Instrumental	APHA (1998)
Gross contamination	Floatables, solvent-soluble floatable oil and grease	APHA (1998)
Chemical		
pH, alkalinity, acidity	Electrometry and titration	APHA (1998), USEPA (1983)
Salinity	Electrical conductivity, density, sensors, titration	APHA (1998), Grasshoff et al. (1999), Parsons et al. (1985)
Dissolved oxygen	Iodometry, oxygen electrode, Winkler method	APHA (1998), USEPA (1983)
Biological oxygen demand	Incubation	APHA (1998), USEPA (1983)
Chemical oxygen demand	Reflux, titrimetry, colorimetry	APHA (1998), USEPA (1983)
Organic carbon	Combustion infrared, persulfate UV oxidation	APHA (1998), USEPA (1983)
Metals	ICPAES, ICPMS, AAS, etc., plus specialist methods for Al, Hg, As, Se, Cr(VI), speciation	APHA (1998), USEPA (1983, 1994c), USEPA (1996c) (Hg); USEPA (1996d) (As)
Ammonia	Ammonia electrode, titrimetry, colorimetry	APHA (1998), USEPA (1983)
Nitrate, nitrite	Colorimetry, ion chromatography, etc.	APHA (1998), USEPA (1983), Grasshoff et al. (1999), Parsons et al. (1985)
Total Kjeldahl nitrogen	Colorimetry, potentiometry	APHA (1998), USEPA (1983)
Carbonate, bicarbonate, CO_2	Titrimetry	APHA (1998), Parsons et al. (1985)
Hardness	Titrimetry	APHA (1998)
Phosphorus	Colorimetry	APHA (1998), USEPA (1983)

Table 5.2. continued

Silica	AAS, Colorimetry, ICPAES	APHA (1998), USEPA (1983), Grasshoff et al. (1999), Parsons et al. (1985)
Cyanide	Titrimetry, colorimetry, cyanide-selective electrode	APHA (1998), USEPA (1993)
Sulfur compounds	Assorted	APHA (1998), USEPA (1983)
Chlorine	Iodometry, amperometry	APHA (1998), USEPA (1983), Grasshoff et al. (1999), Parsons et al. (1985)
Chloride	Colorimetry, titrimetry, IC, potentiometry	APHA (1998), USEPA (1983)
Chlorophyll	Fluorimetry, spectrophotometry	APHA (1998), Parsons et al. (1985)
Oil and grease	Assorted	APHA (1998), USEPA (1983)
Surfactants	Spectrophotometry, etc.	APHA (1998), USEPA (1983)
Phenols	Assorted	APHA (1998), USEPA (1983)
Organochlorine compounds	GC ECD, GC-MS	APHA (1998), USEPA (1983, 1995, 1996e)
Organophosphate pesticides	GC	APHA (1998), USEPA (1983, 1996e)
Carbamate pesticides	HPLC	USEPA (1996e)
Chlorinated phenoxyacid herbicides	GC	APHA (1998), USEPA (1996e)
Dioxins	GC-MS	USEPA (1983, 1996e)
PAHs	HPLC, GC, GC-MS	APHA (1998), USEPA (1996e)
Radioactivity	Counting	APHA (1998)
Biological		
Coliform bacteria	Enumeration, enzyme assays	APHA (1998)
Bioaccumulation	Standard analytical procedures	
Toxicity testing	Bioassays using fish, invertebrates, algae and bacteria	See section 3.5.2

*for APHA (1998), USEPA volumes and AS (1990), use these or the most current editions

Considerable information on the modes of action of the various parameters, their effects on human health, and guidelines on tolerable concentration limits is given in the Water Quality Guidelines (ANZECC & ARMCANZ 2000). The onus is on the practitioner to ensure that environmentally responsible methods are chosen whenever possible. Waste disposal, in particular, must be considered when deciding on a suitable method. It may not be necessary to use an elaborate and expensive method if a more up-to-date, cost-effective and less environmentally damaging technique (such as a field testing procedure) may provide the information at the required level of accuracy.

Before analyses are undertaken, the monitoring team and the end-users of the data should confirm that the chosen laboratory has the appropriate equipment, expertise and experience to undertake the analytical method chosen, as well as an adequate quality assurance program. If the monitoring team plans to send samples to external laboratories, it is recommended that those accredited by the National Association of Testing Authorities (NATA) be used wherever possible. Accreditation guarantees appropriate standards of laboratory organisation and of QA/QC, but not necessarily accurate results.

Results can be reported in a variety of different units such as g/L, ppm, molarity, number of organisms, and so on, but this can cause a great deal of confusion and wasted effort when results from different sources are combined or compared. It is recommended that a system of consistent units is

adopted such as in the Water Quality Guidelines, namely those based on mass/L or mass/kg as appropriate.

5.4. Data Management

When samples are delivered to the laboratory for analysis, it is essential that the laboratory staff log the samples into the laboratory register or record system, and give each a unique identification code. This becomes part of the chain of custody of the sample.

The record system needs to:

- provide a traceable pathway covering all activities from receipt of samples to disposal;
- allow retrieval, for a period of at least three years, of all original test data within the terms of registration.

5.4.1. Data Storage

5.4.1.1. System Design Considerations

Water quality data are expensive to collect, requiring a substantial investment of time and money, so they must made as useable and useful and retrievable as possible by careful systematic storage. The sheer volume of data accumulated by any monitoring program, after just a few years of monitoring, dictates that computer-based data management systems must be the basis for data storage and management.

A data management system should have:

- reliable procedures for recording results of analysis or field observations;
- procedures for systematic screening and validation of data;
- secure storage of information;
- a simple retrieval system;
- simple means of analysing data;
- the flexibility to accommodate additional information, e.g. analytes, sites, etc.

The needs of the user are the most important consideration when a water quality database is being designed. Its designers must consider:

- the scope of the data to be stored — the sources, numbers of samples (and for each one its sample number, type, site, time or date of collection, etc.), number of database fields, descriptive notes, confidence categories, analytes, analysis types, number of records;
- issues caused by multiple sources of data — validation and standardisation procedures, transfer formats, confidence ratings;
- quality assurance and quality control (risk/confidence levels) — analytical precision, validation procedures;
- linkage to flow or tides;
- documentation — standard methods of analysis, validation procedures, codes;
- how the data will be accessed by users — on-line at the same time as it is generated (real-time), on-line retrieval, data retrieval request-based system, categories of data (macro design);
- the kinds of data analysis support required — statistical, graphical, trend analysis, regression analysis.

The Australian and New Zealand Land Information Council (ANZLIC) has a national standard for setting up meta-databases; see their web site, www.anzlic.org.au/asdi/metaelem.htm.

5.4.1.2. Data Tracking

If data are to be the basis for legal proceedings at some time in the future, a chain of custody is particularly important (see also section 4.6.1). In this context, the laboratory may be asked these questions:

- how was each sample labelled to ensure no possibility of mix up or substitution?
- how were the data identified to ensure no mix up or substitution?

The laboratory record system already mentioned will ensure the integrity of the sample from collection to final analysis with respect to the variables of interest. In it, all data need unique identification codes. Then, chain of custody documentation ensures that the questions above can be answered.

5.4.1.3. Screening and Verification

Data entry protocols must be developed to ensure that the entry of data is accurate. Data from instruments should be electronically transferred to the database where possible to prevent transcription errors.

5.4.1.4. Harmonisation of Data

Harmonised data are data that can be used or compared with data from other data sets in comparable units of measurement or time frames. For example, if nutrient loads are to be calculated, concentration data and flow data must be collected at the same location at the same time. To ensure that data are harmonised and can be usefully compared, the monitoring team must consider making additional measurements.

5.4.1.5. Retrieval and Sharing of Data in Databases

A wide range of individual databases has been developed, often associated with the operations systems of particular authorities (water supply, waste water management, storage management). Some of these systems have been costly to update for use with new computer technologies, and have been incompatible with other databases, resulting in difficulties in transferring data.

There has been substantial growth in electronic transfer and on-line access to data in recent years, requiring that databases be standardised. There are a number of commercially available databases, e.g. dB4®, dB5®, ACCESS® and FoxPro®. The adoption of these databases by industry, and the suppliers' commitment to their periodic upgrading to exploit new computer technologies and software developments, ensure that they will continue to have relevance and utility. The choice of a particular database depends on the types and intended uses of the data, and the type and compatibility of the computer hardware and software. The data should be available and able to be shared with other databases for years to come.

5.4.2. Laboratory Data Reporting

Where separate laboratory reports are provided as part of a monitoring program, these should include:

- the laboratory name and address;
- tabulation of samples and analysis data;
- identification of the analytical methods used;
- date of analysis and name of technician or chemist;
- a quality assurance statement.

These details are sometimes reproduced in full, for all relevant samples, in the appendixes of a primary report; otherwise the most appropriate data are abstracted and listed within the body of the report. Further details on reporting are provided in Chapter 7.

5.5. QA/QC in Laboratory Analyses

The objective of a quality assurance and quality control program in a laboratory is to minimise errors that can occur during sub-sampling and analytical measurement and to produce data that are accurate, reliable and acceptable to the data user. Therefore, the QA/QC procedures are designed to prevent, detect and correct problems in the measurement process and to characterise errors statistically, through quality control samples and various checking processes.

5.5.1. Traceability of Results

Traceability of analytical results from the laboratory report back to the original sample is an essential component of good laboratory practice, and is a prerequisite for accreditation of analytical laboratories.

Apart from its chain of custody details for each *sample*, the laboratory record system must include the following information for each *analysis*:

- identity of the sample analysed;
- identity of analyst;
- name of equipment used;
- original data and calculations;
- identification of manual data transfers;
- documentation of standards preparation;
- use of certified calibration solutions.

5.5.2. Laboratory Facilities

It goes without saying that the laboratory facilities and environment must be clean, with appropriate consideration of occupational health and safety issues. In addition, regular checking for airborne contamination is desirable; it can enter through air conditioning systems or be generated internally from users of the laboratory. Deionised water is the most extensively used reagent in the laboratory and it must be maintained at the appropriate standard required to conduct analyses. The electrical conductivity of the deionised water should be monitored continuously or on a daily basis, with the water being checked regularly for trace metals and organic compounds.

5.5.3. Analytical Equipment

All equipment and laboratory instruments should be kept clean and in good working order, with up-to-date records of calibrations and preventative maintenance. Repairs to equipment and instruments should be recorded, as should details of any incidents that may affect the reliability of the equipment.

5.5.4. Human Resources

All staff undertaking analyses must be technically competent, skilled in the particular techniques being used and have a professional attitude towards their work. Thus staff will need to be trained in all aspects of the analyses being undertaken.

Before analysts are permitted to do reportable work, they must demonstrate their competence to undertake laboratory measurements. As a minimum, they should be able to show they can adhere to a written protocol and that their laboratory practices do not contaminate samples. They should demonstrate their ability to work safely in the laboratory and to use the prescribed methods to obtain reproducible data that are of acceptable accuracy and precision.

5.5.5. QA/QC in Analytical Protocols

Laboratories undertaking analyses must fully document the methods used. Methods must be described in sufficient detail that an experienced analyst unfamiliar with a method can reproduce it and obtain acceptable results.

Laboratory staff should be aware that is important to adhere strictly to analytical protocols. They should appreciate the critical relevance of rigorous quality control and assurance in the laboratory. Proper laboratory practice is codified in the requirements of registration authorities such as the National Association of Testing Authorities (NATA). Laboratories holding registration from this and similar organisations will be familiar with the effort required to achieve and maintain a facility with creditable performance standards.

All laboratories must have a formal system of periodically reviewing the technical suitability of analytical methods. If standard methods are used, it is not enough to quote the standard method; any variation of the standard method must be technically justified and supported by a documented study on the effects of the changes.

Measurement errors can be divided into two types: random and systematic.

- *Random errors* affect the precision of the results; that is, the degree to which data generated from repetitive measurements of a sample or samples will differ from one another. Statistically this is expressed as the standard deviation for the replicate measurements of an individual sample and the standard error for replicate measurements of a number of samples. It may also be given as the coefficient of variation; that is, the standard deviation divided by the mean expressed as a percentage. This is often referred to as repeatability or reproducibility. Sources of random error are spurious contamination, electronic noise, uncertainties in pipetting and weighing, etc.
- *Systematic errors* or biases result in differences between the mean and the true value of the analyte of concern (accuracy). Systematic errors can only be established by comparing the results obtained against the known or consensus values. Sources of systematic error are reagent contamination, instrument calibration, method interferences, etc.

The principal indicators of data quality are its *bias* and *precision*. Bias is a measurement of systematic error, and can be attributed either to the method or to the laboratory's use of the method. Precision is the amount of agreement between multiple analyses of a given sample (APHA 1998). When combined, bias and precision are expressed as *accuracy*; that is, the nearness of the mean of a set of measurements to the 'true' value (APHA 1998). The data can be referred to as being accurate when the bias is low and the precision is high (Figure 5.2).

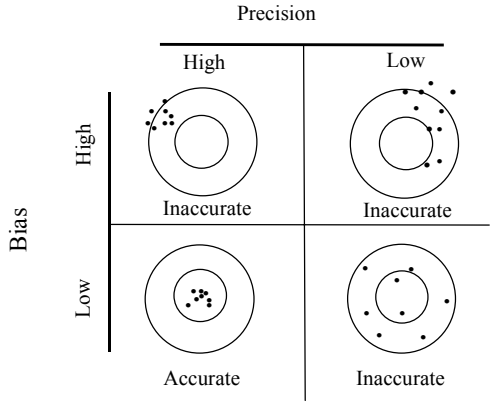


Figure 5.2. Illustration of accuracy in terms of bias and precision (modified from APHA 1998)

Quality assessment is the process of using standard techniques for assessing the accuracy and precision of measurement processes and for detecting contamination. The accuracy of analytical methods, for example, can be established by:

- analysis of reference materials,
- inter-laboratory collaborative testing programs in which it is assumed that the consensus values for analytes are true;
- performance audits;
- independent methods comparisons;
- recovery of known additions;
- calibration check standards;
- blanks;
- replicate analyses.

Similarly the appropriateness of sample storage and preservation procedures can be assessed by inter-laboratory collaborative testing programs. Statistically true values have confidence intervals associated with them, and sample measurements that lie within the confidence intervals will be considered to be accurate.

5.5.5.1. Analysis of Certified Reference Materials and Internal Evaluation Samples

Certified reference materials are materials of known concentrations which have a matrix similar to that of the sample being analysed. The accuracy of laboratory methods and procedures can be established by comparing the values for an analyte in the certified reference material against the results obtained by the laboratory for the same analyte. Results within the confidence limits specified for the certified reference material are deemed acceptable. The National Institute of Science and Technology (USA), Canadian Council for Reference Materials (Canada), International Atomic Energy Commission (Europe), Institute for Reference Materials and Measurements (Belgium) and the National Institute of Environmental Standards (Japan) provide a comprehensive range of certified reference materials.

'Internal evaluation samples' is a general term for samples (made up by an outside source or by the laboratory staff) that have a known analyte concentration, e.g. certified reference materials. The acceptable range of measurement (recovery and precision) is determined for these samples, and analysts are expected to be within this range in all their analyses of these evaluation samples.

5.5.5.2. Proficiency Testing Programs (Interlaboratory Comparisons)

Interlaboratory comparison of unknown samples is used for testing instrument calibration and performance and the skills of the operator. Testing authorities frequently sponsor these programs. Generally only a modest degree of sample preparation is required, probably to restrict the range of sources of variance between laboratories. An individual laboratory compares its results against the consensus values generated by all the laboratories participating in the program, to assess the accuracy of its results and, hence, the laboratory procedures. However, it should be noted that the consensus values can be wrong, and there should be a known value from the authority conducting the proficiency program. Results within the confidence limits specified for the unknown samples are deemed acceptable.

5.5.5.3. Performance Audits

During performance audits, unscheduled checks are made to detect deviations from standard operating procedures and protocols, and to initiate corrective action.

5.5.5.4. Independent Methods Comparison

The accuracy of analytical procedures can be checked by analysing duplicate samples by two or more independent methods. For the methods to be independent they must be based on different principles of analysis. For example, the determination of iron in water can be checked by physical or chemical

principles, e.g. atomic absorption spectrometry (light absorption by atoms) vs. anodic stripping voltametry (electrochemical reduction).

Bias in methods (i.e. interferences, insensitivity to chemical species, etc.) can cause two methods to give different results on duplicate samples. The average values obtained by the methods are compared using a Student's *t*-test (see Appendix section A5.1.6).

5.5.5.5. Recovery of Known Additions

By spiking a test sample with a known amount of analyte, it is possible to estimate the degree of recovery of the analyte and hence the accuracy of the method used. Spiking is one way of detecting loss of analyte. It is assumed that any interference or other effects that bias the method will affect the analyte spike and the analyte in the test sample in similar ways. Hence acceptable recoveries of the spike confirm the accuracy of the method.

This approach may be invalid if:

- the chemical species that are added are different to the native chemical species in the sample and therefore undergo different processes and interferences. An example is the spiking of marine biological samples with AsO_3^{2-}, when the native arsenic is present as arsenobetaine, $CH_3As^+CH_2COO^-$.
- the interference is dependent on the relative concentrations of the analyte and interferent. The addition of a spike will change this dependence and hence the magnitude of the interference.
- the interference is constant, regardless of the analyte concentration. Recoveries can be quantitative but analysis of the native analyte may have large errors.

5.5.5.6. Calibration Check Standards

Standard curves (i.e. calibration curves) must be verified daily by analysing at least one standard within the linear calibration range. This ensures that the instrument is giving the correct response and reduces the likelihood of concentrations in samples being under- or overestimated.

5.5.5.7. Blanks

Blanks should be incorporated at every step of sample processing and analysis. However, only those blanks that have been exposed to the complete sequence of steps within the laboratory will routinely be analysed, unless contamination is detected in them. Blanks incorporated at intermediate steps are retained for diagnostic purposes only, and should be analysed when problems occur, to identify the specific source of contamination.

In principle, only field blanks need to be analysed in the first instance, because they record the integrated effects of all steps. However, a laboratory will normally wish to test the quality of its internal procedures independently of those in the field, so laboratory procedural blanks will usually be included in a suite for analysis, in addition to field blanks.

Blanks cannot be used to detect loss of analyte; they are useful only to detect contamination. They are particularly useful in detecting minor contamination, where the superimposition of a small additional signal on a sample of known concentration may not be evident in the statistical evaluation of analytical data. In other words, blanks are more sensitive to contamination.

If any blank measurement is further than three standard deviations from the mean, or if two out of three successive blanks measurements are further than two standard deviations from the mean, the analyses should be discontinued and the problems identified and corrected.

5.5.5.8. Duplicate Analyses

Duplicate analyses of samples are used for assessing precision. At least 5% of samples should be analysed in duplicate.

5.5.6. QA/QC in Biological Analyses

For biological analyses, quality control procedures are designed to establish an acceptable standard of subsampling, sorting and identification.

5.5.6.1. Subsampling and Sorting

For quality control of subsampling and sorting, a sub-sample equivalent in size to the original subsample should be sent to an independent group, for checks. With macroinvertebrates, the data from the two sub-samples are analysed to compare community composition and structure. The analysis compares the ratios of numbers of taxa in each subsample, and the Bray–Curtis dissimilarities in each subsample. After a sample has been sorted, the remainder is checked for macroinvertebrates that have been missed. Checking continues until >98% of the total number of macroinvertebrates in the sub-sample have been consistently removed.

5.5.6.2. Identification

In general all organisms should be identified against taxonomic keys. If keys are not available, preserved samples should be sent to other laboratories that regularly identify similar samples.

Staff who must identify biological specimens should be trained in the use of keys, and their proficiency should be tested before they are given responsibility for the analysis of samples. For example, in the AUSRIVAS program, new staff identify organisms to family level in samples, and experienced staff check the same samples. The two lists of families are compared and discrepancies are discussed until the new staff understand their errors. The new staff continue to check samples, but fewer samples are cross-checked by the experienced staff as the new staff improve. As many as two samples in 10 may be cross-checked in the early stages of training, but later this drops to two samples in 50. New staff are considered proficient once their error rate at identification to family level is less than 10%.

5.5.7. QA/QC in Ecotoxicity Testing

In ecotoxicity testing, any variability in the test organisms or their health is critical to the quality of the ecotoxicity results. For this reason, standard protocols specifying the life stage and health of an organism are essential. Quality assurance procedures in ecotoxicity tests include criteria for test acceptability, appropriate positive and negative controls, use of reference toxicants, and water quality monitoring throughout the bioassays.

5.5.7.1 Test Acceptability Criteria

All ecotoxicity tests should have criteria for test acceptability. This is particularly important when the test is based on organisms collected in the field, which may vary in their response from season to season. For example, in growth inhibition tests with microalgae, the growth rates of the control group of organisms must exceed a pre-defined daily doubling rate with less than 20% variability. Similarly, in acute tests with invertebrates and fish, at least 90% of the untreated control organisms must be alive after 96 hours. Fertilisation or reproduction tests usually specify an acceptable fertilisation rate, abnormality rate or number of offspring produced. If these criteria are not met, the tests are invalid and must be repeated.

5.5.7.2. Negative Controls

All toxicity tests require the use of controls to compare the responses of the organisms in the presence or absence of toxicant. Negative controls can be uncontaminated seawater or freshwater used as diluent in the toxicity tests, with water quality characteristics similar to those of the test water. For sediment tests, negative controls include uncontaminated sediments that have particle sizes, organic carbon and sulfide contents similar to those of the test sediment.

5.5.7.3. Reference Toxicants

Reference toxicants or positive controls are used to ensure that the organism on which the toxicity test is based is responding to a known contaminant in a reproducible way. This is particularly important for field-collected organisms, which may vary in response to a toxicant depending on season, collection site, temperature and handling.

Reference toxicants are also used to track changes in sensitivity of laboratory-reared or cultured organisms over time. Usually, either inorganic (e.g. copper, chromium) or organic (e.g. phenol, sodium dodecyl sulfate) reference toxicants are used and tested at a range of concentrations on a regular basis. In addition, each toxicity test should include at least one concentration of reference toxicant as a positive control. Quality control charts are produced showing the mean response and variability over time. Further guidance on the use of reference toxicant tests can be obtained from Environment Canada (1990b) and USEPA (1993).

5.5.7.4. Blanks

Appropriate field and process blanks should be included in each toxicity test if the sample has been manipulated before testing (see also section 4.6.3.1 in Chapter 4). If freshwaters have to be salinity adjusted with artificial sea salts before testing, sea salt controls should also be included in each test. Solvent controls are also essential for water insoluble chemicals if they have to be dissolved in organic solvents to deliver them into the test system. It is important to test a solvent control with the same concentration of solvent in clean water as is found in the highest test concentration. In no case should the concentration of solvents or emulsifiers exceed 0.1 mg/L (OECD 1981).

5.5.7.5. Quality of Ambient Water

Throughout the toxicity tests, the quality of the organisms' ambient water must be monitored to ensure that the toxicity measured is due to the contaminant or test sample alone, and to provide information that can be used in test interpretation. For freshwaters, measurements of alkalinity, hardness, pH, temperature and dissolved oxygen are the minimum parameters required. For marine studies, salinity is also monitored throughout the test.

5.5.8. QA/QC for Handling Sediments

The principles of handling sediment samples are similar to those for water analyses. Sample integrity must be maintained and QA/QC are important, as described in section 5.5.5. Some other aspects of the handling of sediment samples have been touched on already: the need to maintain unchanged redox conditions (oxic vs. anoxic) in the samples (section 4.3.5); and the need for negative controls and test samples to have similar particle sizes, organic carbon and sulfide contents (section 5.5.7.2). For more detailed advice on sediment handling and preparation, see Mudroch and Macknight (1991). Sediment sampling, handling and analysis are also described more fully in the Water Quality Guidelines, Volume 2, Appendix 8 (ANZECC & ARMCANZ 2000).

5.5.8.1 Pore Water Sampling

Pore waters in sediments can be sampled by pressure filtration or by centrifugation or by in situ methods such as pore water dialysis cells (peepers), gel samplers, or sippers. All operations should be conducted in an inert atmosphere. Filtration and centrifugation are the techniques most commonly used, but their suitability will depend on the sediment grain size. Solvent displacement has also been successfully applied in centrifugation methods.

The use of sippers or direct withdrawal techniques is limited to sandy sediments with a large pore volume. For investigation of pore water depth profiles, the peeper and gel sampling methods offer the best option.

5.5.8.2 Sample Storage

Sediment samples are typically either chilled or frozen immediately after collection, for storage and to minimise bacterial activity. Where total contaminant concentrations are of interest, either of these methods is suitable. In addition, oven drying at 110°C is also an option. For most organic compounds and the more volatile metals, e.g. Hg, Cd, Se, As, oven drying is unacceptable, and air drying at room temperature or freeze drying is preferable, although even these may be a concern with very volatile organic compounds.

Special considerations are required where metal speciation is of concern or where pore waters are to be analysed. Since most sediments are anoxic, at least in part, oxidation of iron sulfides in particular will change the chemical forms of metals in the pore water and solid sediment phases. This oxidation can be minimised by freezing the sample and storing it in a sealed container, and by preferably carrying out such operations in an inert atmosphere, e.g. under a nitrogen gas blanket or in a glove box. Freeze-drying can be done in an oxygen-free environment such as a glove box.

Freezing can rupture biological cells and release metals, and, although this is not generally an issue for sediments, it can significantly bias pore water results for some elements, e.g. selenium.

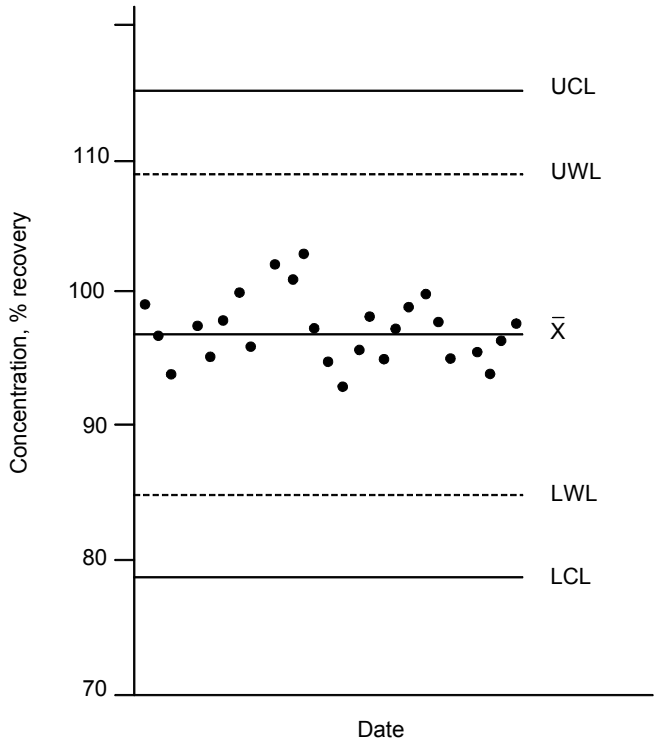


Figure 5.3. An example of a control chart for mean values; UCL = upper control limit; LCL = lower control limit; UWL = upper warning limit; LWL = lower warning limit

5.5.8.3. Sieving Samples

Sieving is the process used to divide sediment samples into fractions of different particle size. Sediments are usually classified as gravel (>2 mm), sand (63 µm to <2 mm), silt and clay (<63 µm). Sediments are usually sieved through a series of mesh sizes from 2 mm to 63 µm. Wet sieving is used for processing fine grain sediments while sieving of dry material is used for the separation of coarser material. When comparing trace metal concentrations in sediments from different sampling sites, it is normal to analyse the <63 µm fraction because it is this fraction that adsorbs most of the trace metals.

It should be noted that if organic contaminant concentrations in sediments are to be compared, grain size is not important but concentrations should be expressed as a percentage of organic carbon content.

5.5.8.4. Homogenisation of Samples

It is difficult to ensure the homogeneity of sediment samples being analysed because samples are notoriously heterogeneous with respect not only to particle size but also to contaminant distribution. Thorough mixing of any wet or dried samples is required to improve homogeneity. For dry samples, grinding with a mortar and pestle is necessary. Commercially available rock grinders are used to reduce larger dried particles to less than 63 µm for analysis. Coning and quartering, rolling, mechanical mixing and splitting are used for homogenisation and selection of material for analysis (Mudroch and MacKnight 1991).

Wet samples are used where it is feared that drying will alter the chemical form of the contaminants. For large wet sample volumes, homogenisation is especially difficult. It is more usual to use wet samples with smaller volumes, e.g. core sections. Here the wet samples can be homogenised by thorough mixing with a glass rod, then weighed out for analysis, with moisture determinations being carried out on separate aliquots of the sample.

5.5.9. Presentation of Quality Control Data

Control charts are used to visualise and monitor the variability in QC data (Lewis 1988). Two types of control charts are commonly used in laboratories: means charts and range charts.

- *Means charts* are used to track changes in certified reference concentrations, known additions, calibration check standards and blanks. The charts are graphs of the mean ± standard deviation or error over time (Figure 5.3) with a defined upper and lower control limit (normally three times the standard deviation where 99.7% of the data should lie). These are the limits. Data that fall above or below these limits are unacceptable and corrective action must be taken. Normally, action is taken if data are trending towards these limits.
- *Range charts* are used to track differences between duplicate analyses based on the standard deviation or relative standard deviation. Again, limits are set; when data fall above or below these limits, corrective action must be taken. For a full explanation of the procedures for calculating ranges refer to APHA (1998 or most current edition).

The number of quality control samples will depend on two considerations: the duration of acceptable deviations from the mean that can be tolerated, and the likelihood of deviations occurring. The first consideration will depend on the purpose for which the data are being collected, and the number of quality control samples will be set in consultation with the user of the data. The second consideration is usually based on past variability. To be effective, control charts must be continually updated as data become available so trends can be established before control limits are reached.

As a minimum, the precision and accuracy of data must be stated when data are presented in reports. Precision, as standard deviation or error, or relative standard deviation or error, should be presented graphically across the range of values being measured (Figure 5.4a). Alternatively, the standard deviation or similar should be given, at discrete values that cover the range of measurements. The method(s) of establishing the accuracy of the data should be stated. Bias should be presented graphically as a function of the true value over the applicable range (Figure 5.4b). Alternatively the bias at discrete values that cover the range of measurements should be given.

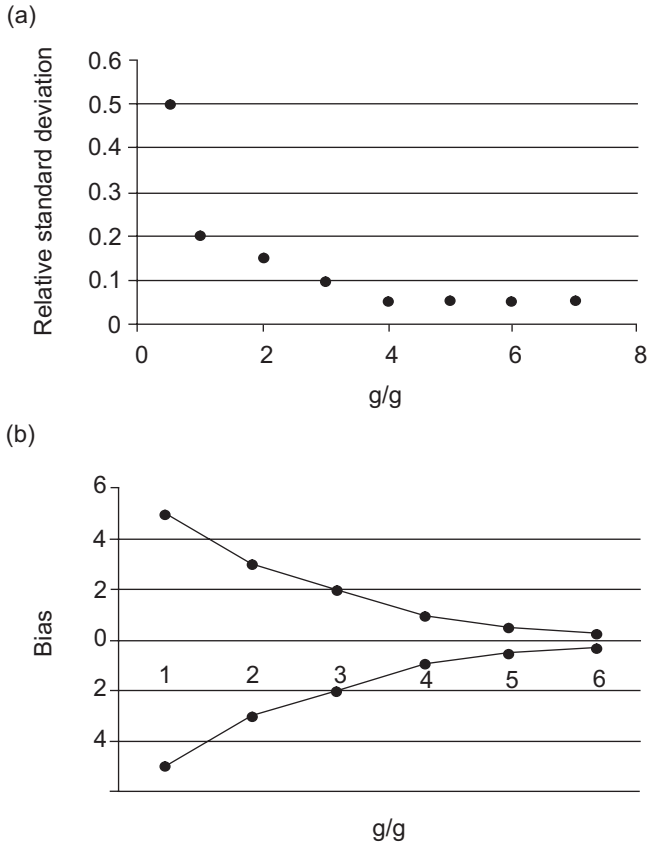


Figure 5.4. Graphical representation of (a) precision; (b) bias

5.6. Occupational Health and Safety

5.6.1. Legislative Requirements

Occupational health and safety requirements for laboratory work are provided in the various parts of Australian/New Zealand Standard AS/NZS 2243.2 (e.g. AS/NZS 1995, 1997a,b), *Safety in Laboratories*. This standard sets out the recommended procedures for safe working practices in laboratories.

Aspects covered include:

1. general;
2. chemical aspects;
3. microbiology;
4. ionizing radiators;
5. non-ionizing radiators;
6. mechanical aspects;
7. electrical aspects;
8. fume cupboards;
9. recirculating fume cupboards.

Practical guidance to the safety procedures and information needed for scientific work and safe laboratory practice is also available in *Safety Manual: An Essential Reference for Every Laboratory* (Haski et al. 1997).

5.6.2. Identification of Hazards

The hazards or risks involved with laboratory work need to be identified and documented. The major issues are:

- whether or not staff will be exposed to toxic or other hazardous substances;
- whether or not staff will be placed in a position of potential physical danger.

Staff who are to conduct analyses should be physically and mentally able to carry out laboratory work.

5.6.3. Education about Hazards

All staff must be appropriately trained as part of the formal risk minimisation strategy. Training will include:

- familiarisation with protocols (analysis procedures, safe handling procedures, disposal procedures, chain of custody considerations, etc.);
- use of laboratory equipment;
- qualifications in handling chemicals;
- familiarisation with safety procedures;
- qualifications in advanced first aid.

5.6.4. Risk Minimisation Plans

Proper professional practice requires that risks be reduced as much as possible, and that staff do not have to work in unsafe conditions.

Actions that can be taken to reduce risks include:

- the wearing of appropriate clothing and footwear to protect against accidental chemical spills;
- provision of an appropriate first aid kit in close proximity to where analyses are being undertaken;
- provision of an eye bath and safety shower in the laboratory;
- the training of laboratory staff in first aid procedures;
- staying in contact with help and never working alone; that is, at least three staff should work together and be in contact with someone who can raise an alarm. There should be written procedures describing how emergency services are to be contacted.

Chapter Six
Data Analysis and Interpretation

6.1. Introduction

This chapter provides guidance on the use of common statistical methods for the analysis of water quality data. The information is at an introductory level that will help the monitoring team identify suitable methods of analysis and assist with the interpretation of results. Much of the technical detail and the more advanced statistical procedures have been relegated to Appendix 5 where, in most cases, they are illustrated with the help of worked examples that demonstrate options available, methods of implementation and interpretation of results. The information provided in this chapter is not exhaustive; complex studies may require a greater level of statistical sophistication than is presented here, and for these studies the monitoring team is advised to consult a professional statistician.

Data analysis should be viewed as an integral component of the water quality management process. A framework for undertaking data analysis and interpretation is shown in Figure 6.1. A checklist is presented in Table 6.1.

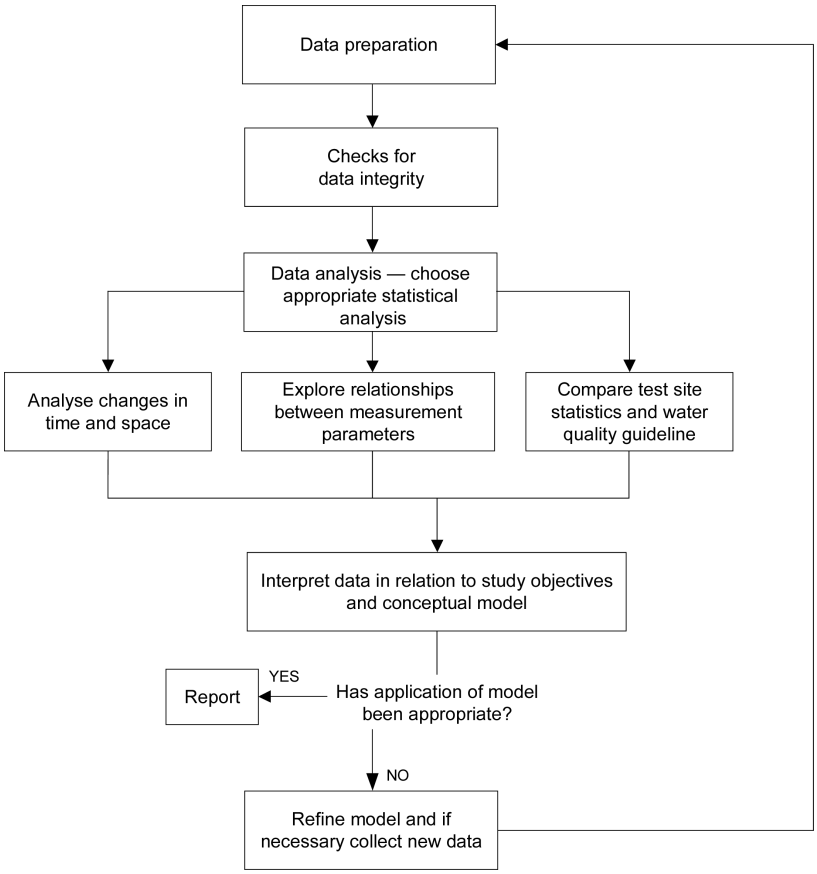


Figure 6.1. Framework for data analysis and interpretation

Table 6.1. Checklist for data analysis

1.	Preliminaries before data analysis; data preparation
	(a) Has the purpose of the data analysis exercise been clearly stated?
	(b) Have the parameters to be estimated and/or hypotheses to be tested been identified?
	(c) Will data from different sources be compatible (levels of measurement, spatial scale, time-scale)?
	(d) Have objectives concerning quality and quantity of data been articulated?
	(e) Has a program of statistical analyses been identified?
	(f) Have the assumptions that need to be met for correct application of statistical methods been listed?
	(g) Has data organisation (storage media, layout, treatment of inconsistencies, outliers, missing observations and below detection limit data) been considered?
2.	Have data reduction methods (graphical, numerical, and tabular summaries) been applied?
3.	Have 'aberrant' observations been identified and remedial action taken?
4.	Have potential violations of statistical assumptions (e.g. non-normality, non-constant variance, autocorrelation) been checked for?
5.	Have data been suitably transformed if necessary?
6.	Have data been analysed using previously identified methods; have alternative procedures been identified for data not amenable to particular techniques?
7.	Have results of analysis been collated into a concise (statistical) summary and have statistical diagnostics (e.g. residual checking) been used to confirm the utility of the approach?
8.	Has the statistical output been carefully assessed and given a non-technical interpretation?
9.	Have the objectives been addressed? If not, has the study been redesigned, have new or additional data been collected, the conceptual models refined and data re-analysed?

As noted in earlier chapters, data types, quantities, and methods of statistical analysis need to be considered collectively and at the early planning stages of any monitoring strategy. Decisions must be made about:

- measurement scales,
- frequency of data collection,
- level of replication, and
- spatial and temporal coverage,

so that data of sufficient quality and quantity are collected for subsequent statistical analysis. It is also important for the monitoring team to avoid the 'data rich–information poor' syndrome of collecting data that will not be subsequently analysed or that do not address the monitoring objectives.

Given the costs associated with the data collection process, it is imperative for the monitoring team to use formal quality assurance and quality control (QA/QC) procedures to ensure the integrity of the data. These procedures should be supported by established back-up and archival processes. There should be serious consideration of the archival medium to be used, because rapid advances in computer technology tend to increase the speed at which equipment becomes obsolete. The quality assurance of any monitoring program should include the data analysis, as well as the field and laboratory practices. Before statistically analysing the monitoring data, the monitoring team should use standard methods of data summary, presentation, and outlier checking to help identify 'aberrant' observations. If undetected, these data values can have profound effects on subsequent statistical analyses and can lead to incorrect conclusions and flawed decision making.

The monitoring team needs to develop a plan of the sequence of actions they will use for the statistical analysis of the water quality data. Only some of the many available statistical techniques need be used. An initial focus of monitoring might be to assess water quality against a guideline value or to detect trends, but an ultimate objective of the data analysis exercise will probably be to increase

the team's understanding of the natural system under investigation. Improved understanding should result in more informed decision making which in turn will lead to better environmental management. One of the most significant challenges for the data analysis phase is to extract a 'signal' from an inherently noisy environment.

As discussed in section 3.2 (Chapter 3), monitoring study designs will fall into three basic categories: descriptive studies including audit monitoring; studies for the measurement of change, including assessment of water quality against a guideline value (although that sort of study can also be categorised as descriptive); and studies for system understanding, including cause-and-effect studies and general investigations of environmental processes. The statistical requirements for the descriptive studies are less complex than for the other categories in which more detailed inferences are being sought.

Most of the statistical methods presented in this chapter are based on classical tools of statistical inference (e.g. analysis of variance, *t*-tests, *F* tests, etc.). These methods have served researchers from a variety of backgrounds and disciplines extremely well over many years. However, there is growing concern about their utility for environmental sampling and assessment. When natural ecosystems or processes are being measured it is invariably hard to justify the assumptions that response variables are normally distributed, that variance is constant in space and time, and that observations are uncorrelated. In these cases, remedial action (e.g. data transformations) may overcome some of the difficulties, but it is more probable that an alternative statistical approach is needed. For example, generalised linear models are often more suitable than classical ANOVA techniques for the analysis of count data because of their inherent recognition and treatment of a non-normal response variable. Recently, a number of researchers have questioned the utility and appropriateness of statistical significance testing for environmental assessment (e.g. McBride et al. 1993; Johnson 1999; and see section 6.4.2). Also, there is a progressive move towards more 'risk-based' methods of water quality assessment; although some of these methods also have difficulties (Fox 1999).

A large number of introductory and advanced texts of statistical methods are available (e.g. Ott 1984; Helsel and Hirsch 2000). For a relatively detailed and comprehensive description of statistical techniques used in water quality management studies, see Helsel and Hirsch (2000).

Statistical formulae are presented in this chapter and in Appendix 5 for clarification or for completeness, but it is recognised that the monitoring team will consign most computation to a statistical software package.

There is a plethora of statistical software tools on the market and it is beyond the scope of the Monitoring Guidelines to review them all. However, several are worthy of mention. SAS® is a powerful data analysis tool suited to handling large data sets; MINITAB®, STATISTICA® and SYSTAT® are suited to the analysis of medium–large data sets and their popularity stems from ease of use and a comprehensive set of procedures; S-PLUS® embodies many contemporary and robust statistical procedures which are not readily available elsewhere. Microsoft EXCEL® is useful for data screening and manipulation. EXCEL® has a limited number of intrinsic statistical functions, and a number of third-party statistical 'add-ins' are also available. Concerns have been raised about the accuracy and numerical stability of the statistical algorithms in EXCEL®; based on the results of rigorous testing, McCullough and Wilson (1999) concluded that 'persons desiring to conduct statistical analyses of data are advised not to use EXCEL'. However, while this is important when accuracy to many decimal places is needed or when algorithms must deal with 'ill-conditioned' data sets, it is unlikely to have measurable effects on the types of analyses contemplated by water quality scientists and natural resource managers.

All these software tools provide a high level of functionality and technical sophistication, but they also lend themselves to abuse through blind application. The monitoring team should not indulge in unstructured and undisciplined application of techniques because they 'seem to work' or produce a predetermined outcome. However, the team may find Exploratory Data Analysis (EDA) and data mining useful.

The remainder of this chapter outlines the data analyses ordinarily undertaken in practice. The first part recommends and suggests techniques for summarising data sets and reducing them to descriptive numerical quantities, and for data visualisation, transformations, outlier detection, censoring, trend detection and smoothing. Statistical techniques referred to here are designed to help tease out patterns in data sets, identify anomalies, and condense raw data using graphical tools and summary statistics. Later sections of the chapter cover a wide range of 'classical' statistical methods for helping the monitoring team make decisions about the likely values of water quality parameters, observe changes in these parameters over space and time, and explore relationships between pairs or groups of parameters. The later sections also refer to contemporary statistical techniques such as generalised additive models, robust regression and smoothing; these are usually computationally intensive and can only be realistically undertaken with a fast computer and appropriate statistical software.

6.2. Data Preparation

The data obtained from laboratory and field studies need to be summarised in a form that is amenable to analysis (see also section A5.1.1). It is best to check that the data provided are only those that are acceptable by field and laboratory quality assurance/quality control (QA/QC) criteria, and that they are rounded off to the appropriate number of significant figures.

Analytical data can be tabulated into spreadsheets that are immediately useable for exploratory analysis. Physical measurements should be tabulated in a form that permits ready comparisons with chemical and biological data for the same sites. The choice of formats for these is reasonably intuitive.

Graphical presentations of the basic data are also useful for displaying differences; e.g. a profile can illustrate changes in metal concentrations with depth in a sediment core.

Prior to more comprehensive analysis, the data must be given a preliminary examination to check their integrity before they can be subjected to more detailed analysis. Data that are missing or below detection limits (so-called 'censored' data) will need to be considered, as will obvious outliers that might be attributable to experimental or transcription errors.

6.2.1. Censored Data

Unless the water body is degraded, failure to detect contaminants is common. Rather than concluding that the particular contaminant does not exist in a water sample, the monitoring team records the observation as 'below detection limit' (BDL). Unfortunately, there is no universally accepted method of dealing with BDL data. Some common approaches include these:

- treat the observation as 'missing';
- treat the observation as zero;
- use the numerical value of the detection limit;
- use the numerical value of half the detection limit.

When a large portion of the data is below detection limit, use of any of the above approaches will be problematic because the sample variance will be severely underestimated. Also, when standard statistical techniques are applied to data sets that have constant values in place of the BDL values, the resulting estimates are biased (El-Shaarawi and Esterby 1992).

Assigning a missing value code (such as * or NA) can also cause difficulties because the various software tools treat missing values differently. For example, some software packages compute an average using the full data set with missing values replaced by zeros (not a good strategy) while others simply ignore all missing values and reduce the sample size accordingly. More sophisticated statistical procedures have been devised that use re-sampling or imputation techniques to 'infer' a reasonable value for the BDL observation; their implementation requires advanced statistical skills.

Gilliom and Helsel (1986) and Helsel and Gilliom (1986) have estimated water quality parameters with censored data using a variety of techniques, while Liu et al. (1997) describe a method based on 'forward censored regression' to model data with censored observations. Commercial software packages for 'data imputation' have recently become available, although their narrow focus and restricted audience will probably not have any measurable impact on the way environmental managers treat BDL data.

In the absence of more sophisticated tools for analysing censored data it is suggested that routine water quality parameters (means, percentiles, etc.) be computed using the full data set with BDL data replaced by either the detection limit or half the detection limit. The impact of this strategy on computed statistical measures should be clearly understood, and the monitoring team should *not* proceed with any form of inferential analysis (e.g. confidence intervals or hypothesis testing) when a significant proportion (e.g. >25%) of the data set is BDL and has been substituted by a surrogate value. Advanced statistical skills should be sought when any form of inference is required in these situations.

If only a small proportion of the data set is BDL and has been replaced by a numerical surrogate, it is best to perform any statistical analysis twice, once using zero and once using the detection limit (or half the detection limit) as the replacement value. If results from the two analyses differ markedly, the monitoring team should investigate more sophisticated statistical methods of dealing with censored observations (e.g. software packages LIMDEP by Econometric Software and SHAZAM by University of British Columbia). If the results do not differ markedly, the censored observations probably have little influence on the analysis.

6.2.2. Data Integrity

The integrity of water quality data can be reduced in many and varied ways. Losses or errors can occur at the time of collection, and in the laboratory during sample preparation and analysis, and during recording of results, and during electronic manipulation and processing, and during analysis, interpretation and reporting. Once the 'certified' data leave the laboratory, there is ample opportunity for 'contamination' of results to occur. Gross errors that are probably the result of data manipulations (transcribing, transposing rows and columns, editing, recoding, and conversion of units) are easily overlooked unless a modest level of screening is undertaken. While these sorts of errors can usually be detected by a scan of the raw data, more subtle effects (e.g. repeated data, accidental deletion of one or two observations, or mixed scales) are more difficult to identify. If left unchecked, these 'anomalous' readings can have a profound impact on subsequent statistical analyses and possibly lead to erroneous conclusions being drawn.

A good QA/QC program for data analysis uses simple yet effective statistical tools for screening data as they are received from laboratories. These typically include a mixture of graphical procedures (histograms, box-plots, time sequence plots, control charts, etc.) and descriptive numerical measures of key distributional aspects (mean, standard deviation, coefficient of variation, and possibly measures of skewness and kurtosis — see Appendix 5 (e.g. section A5.1.1) for a more complete description of these). These analyses are routine, but should not be ignored. Neither should they be unsupervised. The data analyst should oversee all processing of the data and carefully inspect graphs and reports. Most if not all pre-processing, manipulation and screening can be undertaken using the built-in capabilities of an electronic database system. The treatment of 'unusual' observations or 'outliers' is more contentious and is discussed next.

Care should be exercised in labelling extreme observations as 'outliers'. An outlier is indeed an extreme observation, although the converse is not necessarily true. It should be kept in mind that about two out of every 1000 observations from a normal distribution will fall beyond three standard deviations from the mean. To automatically label these as outliers and discard them from the data set would introduce bias into subsequent analyses. On the balance of probabilities, an observation beyond three standard deviations from the mean is likely to be 'aberrant'. Such observations need to be

highlighted for follow-up investigation to identify causes (e.g. recording error, laboratory error, abnormal physical conditions during sampling).

If there is no rational explanation for its value, the decision to include or exclude an outlier from the data set rests with the data analyst. It is suggested that only the most extreme observations (e.g. those that are four or more standard deviations from the mean) be excluded unless other good reasons can be established (e.g. different analytical method used). There are statistical tests for determining if a specific value can be treated as an outlier (see Neter et al. 1996) and techniques such as the box-plot can help in the decision. However outliers may convey significant information and their presence should initiate more thorough investigation[1]. Simple descriptive statistical measures and graphical techniques, combined with the monitoring team's knowledge of the system under investigation, are very valuable tools for identifying outliers.

Identification of aberrant observations in a multivariate (i.e. many variable) context is more complex. Water quality monitoring generates measurements of, say, metals, nutrients, organic compounds and other compounds, so for each sample there is often a vector of observations rather than a single value. These values (variables) tend to be correlated among themselves to some extent, indicating that they co-vary. If the co-dependence between variables is ignored and the aberrant observations are examined one variable at a time, unusual observations may be missed, and the whole procedure will be inefficient. In a multivariate context, it is quite possible for an observation to be 'unusual' even when it is reasonably close to the respective means of each of the constituent variables.

It is even more difficult to determine the causes of outliers in a multivariate context. Garner et al. (1991) have considered statistical procedures to help in this evaluation, but again care needs to be exercised in the labelling and treatment of potential outlying observations. The issue of outliers and multivariate normality for compositional data (i.e. where data represent percentages of various chemical constituents) has been investigated by Barceló et al. (1996). See also section A5.1.3.

6.3. Exploratory Data Analysis (EDA)

The way in which data are analysed will largely depend on the type of study being undertaken. The framework in Figure 6.1 suggests paths for analysing data from descriptive studies or from studies that measure change or system understanding; these will largely have been decided in the study design process, which defined the particular questions that need to be addressed (Chapter 3). This section discusses the range of data processing tools that might now be used.

6.3.1. Data Reduction

Data reduction is an important first step in the data analysis process. It enables the monitoring team to present and summarise important features of the data, and it also helps the analyst identify outliers or observations that are 'aberrant' in some other way.

A combination of statistical tools can be used for data reduction, including graphs (e.g. histograms, box-plots, dot-plots, scatterplots), tables (e.g. frequency distributions, cross-tabulations), and numerical measures (e.g. means, medians, standard deviations, percentiles). The objective of calculating summary statistics is to convey the essential information contained in a data set as concisely and as clearly as possible and/or to estimate a parameter of some population of values (see also section A5.1.1).

Frequency tabulations are commonly used for summarising data because they can condense even large data sets to manageable form without substantial loss of information, and because they can be graphically presented in the form of histograms and bar-charts.

[1] There is a salutary lesson to be learned from the automatic discarding of outliers. The hole in the ozone layer went initially undetected because computers used in the processing of the data had been programmed to remove extreme observations prior to analysis.

It is common to use statistics that are measures of the central tendency or 'average', such as the arithmetic mean, the median and the mode. A description of some common 'averages' is given in Table 6.2. For most water quality applications the arithmetic mean, geometric mean or median are the most appropriate quantities to use. The median is a robust estimator of central tendency because it is relatively unaffected by extremes in the data. The arithmetic mean does not share this property, but it is nevertheless the most widely used 'average': it is easy to compute and uses all the data. Also, it has well-established statistical properties, and many statistical tests relate to inference about a population mean. However, the choice of measure should not be automatic and will depend on the circumstances at hand. The arithmetic mean is often appropriate for the computation of loads of a contaminant, while a median or geometric mean is the preferred statistic for describing an 'average' concentration. For the arithmetic, geometric, and harmonic means the following is always true:

$$\text{harmonic mean} \leq \text{geometric mean} \leq \text{arithmetic mean}.$$

The three means are only equal if all the sample values are identical. Parkhurst (1998) discusses in more detail the relative merits of the geometric mean and arithmetic mean in the context of concentration data.

Variability is another extremely important characteristic of a distribution of results (see Table 6.3). The simplest measure of variability is the range — the difference between the largest score and the smallest score. The range is rarely used as a measure of variability because it is grossly affected by extremes in the data set (after all, it is defined as the difference between the two extremes). The most commonly used measure of variability is the variance or its (positive) square root, the standard deviation. The standard deviation is preferred because it has the same units of measurement as the original data. However, the variance is an important parameter in inferential statistics such as analysis of variance. One difficulty with the standard deviation is that it is not readily comparable among different populations or samples because it tends to be numerically higher as a result of an increasing mean. A dimensionless measure that overcomes this difficulty is the coefficient of variation (CV) which is defined as the ratio of the standard deviation and the mean.

Table 6.2. Measures of central tendency (adapted from Spiegel 1992)

Arithmetic mean	$\bar{X} = \sum_{i=1}^{n} X_i / n$ where X_i denotes the ith observation in a sample of n
α% trimmed mean	$\bar{X}_{T,\alpha}$ obtained by trimming (i.e. removing) α% off both ends of the ordered sample and computing $\bar{X}$ for the remainder. Used when outliers are present. α is typically 10% or 20%.
Mode	most frequently occurring value
Median	the middle value: half the values are numerically smaller and half are numerically larger
Geometric mean	the nth root of the product of n sample values (>0): GM = $(x_1 x_2 \ldots x_n)^{1/n}$. It is always less than the mean.
Harmonic mean	the reciprocal of the summation of n sample reciprocal values: HM = $\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}$

In addition to measures of location and dispersion (variability), another statistical quantity that has assumed greater significance and utility in water quality applications is the percentile of a distribution. For example, when examining physical–chemical stressors, the Water Quality Guidelines (ANZECC & ARMCANZ 2000) have adopted a triggering process based on a comparison of the 50th percentile at a test site with the 80th percentile at a reference site.

The pth percentile is the value that is greater than or equal to $p\%$ of all values in a distribution; e.g. 50% of all values in a distribution are numerically less than or equal to the 50th percentile (otherwise known as the median) while 80% of all values are numerically less than or equal to the 80th percentile. The 25th, 50th, and 75th percentiles are called the quartiles (denoted Q_1, Q_2 and Q_3) because they divide the distribution into four parts of equal probability.

Table 6.3. Common measures of variation (adapted from Spiegel 1992)

Range	(largest value) – (smallest value)
Interquartile range (IQR)	75th percentile – 25th percentile (percentiles and quartiles are defined in the text)
Sample variance (s^2)	$s^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 / (n-1)$
Sample standard deviation (s)	The square root of the variance; it has the same units as the central tendency statistics
Windsorized standard deviation (s_T)	Used as a measure of spread of the trimmed mean $\bar{X}_T$ (Table 6.2). Obtained by replacing trimmed values, identified in computation of $\bar{X}_T$, with values that were next in line for trimming (one from each end), and computing the standard deviation s of this new sample; s_T is then obtained as $s_T = s\sqrt{(n-1)/(k-1)}$ where k is the size of the trimmed sample.
$\frac{\text{IQR}}{\text{median}}$	A dimensionless robust measure of spread
Coefficient of variation (CV)	The standard deviation divided by the mean; it is therefore dimensionless and can be used for comparisons among different samples

Table 6.4. Taxonomy of common types of graph and their applications

Graph type	EDA	Outlier detection	Distribution checking	Trend detection	Multivariable relationships	Model diagnostics	Process control	Cause and effect
Scatterplot (bi-plot)	✓	✓		✓	✓	✓		
Scatterplot matrix	✓	✓		✓	✓	✓		
Trellis graph	✓				✓			
Chart (line, bar, pie)	✓			✓				
Interval (error bars, confidence intervals)		✓				✓		
Histogram	✓	✓	✓			✓	✓	
Box-plot	✓	✓	✓			✓	✓	
Dot-plot	✓			✓				
Stem and leaf	✓	✓	✓			✓		
Time series	✓	✓		✓		✓	✓	
3D (contour, scatter, surface)	✓	✓		✓	✓			
Probability and Q-Q plots	✓	✓	✓					
Scatterplot smoothers (LOESS)	✓			✓				
Residual plot		✓	✓	✓		✓		
Run chart	✓			✓			✓	
Control chart (Xbar, CUSUM, EWMA, Range, MA)	✓						✓	
Pareto	✓						✓	
Fishbone	✓							✓

6.3.2. Data Visualisation

With recent advances in computer hardware and software, high quality and sophisticated graphics have become readily accessible. It is strongly recommended that relevant graphs of the data are drawn before any formal statistical treatment and analysis are done (Tufte 1983). Simple graphical devices such as histograms, box-plots, dot-plots, scatter plots, probability plots, etc., can assist statistical analyses and make it easier to interpret data with respect to:

- data anomalies and errors,
- outliers,
- properties of the distributions (location, dispersion, skewness),
- trends over time, space, attributes,
- relationships (existence of and type),
- checking assumptions appropriate to the distributions (e.g. normal probability plots),
- time series analysis,
- reducing the number of dimensions (visualising high dimensional data by projecting it into lower dimensions),
- operational performance (e.g. control charts).

A list of graph types and potential applications is given in Table 6.4. Most of these plots are available in the statistical software packages referred to in section 6.1.

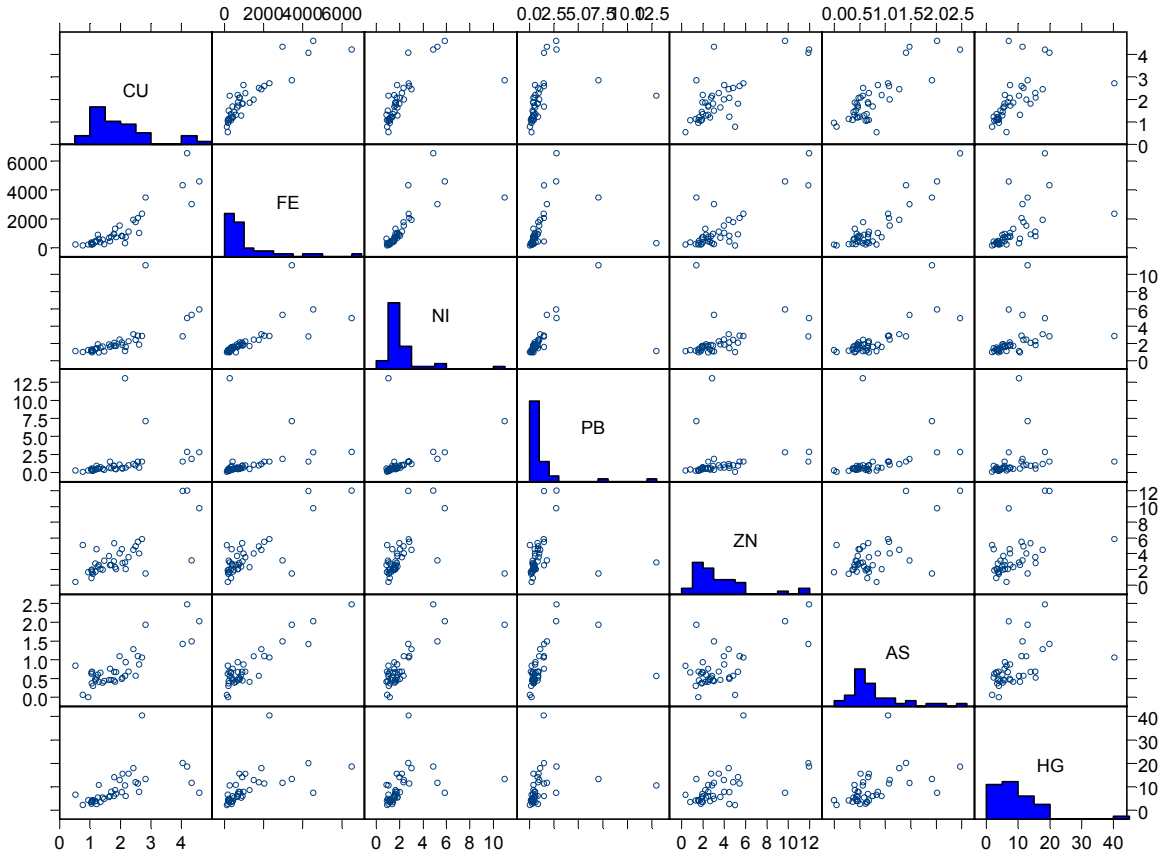


Figure 6.2. Scatterplot matrix for concentration of seven metals from a single water sample

For most water quality analyses, several different attributes or measures of water quality have to be assessed simultaneously. Important information can be overlooked if water quality variables are analysed one at a time. Relationships between individual variables can be detected relatively easily from graphical summaries such as the scatterplot matrix in Figure 6.2, which shows scatterplots and histograms for the concentrations of seven metals in water samples from a Victorian lake. A graphical display like this enables the analyst to assimilate the trends, relationships, and distributions of a number of variables that have been measured on the same water sample. Histograms for individual variables are shown on the diagonal while plots for pairs of variables are displayed on the off-diagonal (the labelling of axes is reversed to create two plots for each pair of variables on the off-diagonal).

6.3.3. Control Charting

Statistical Process Control (SPC) dates back to the 1930s and is deeply rooted in industrial applications where it is vitally important to control drift and variation in a process, to maintain production quality. Control charting techniques used for the last 70 years in industry have an important role to play in an environmental context. They are particularly relevant to water quality monitoring and assessment. Regulatory agencies are moving away from the 'command and control' mode of water quality monitoring, and recognising that, in monitoring, the data generated from environmental sampling are inherently 'noisy'. The data's occasional excursion beyond a notional guideline value may be a chance occurrence or may indicate a potential problem. This is precisely the situation that control charts target. They not only provide a visual display of an evolving process, but also offer 'early warning' of a shift in the process level (mean) or dispersion (variability). For further information, see Montgomery (1985) or Bennett and Franklin (1988 Chapter 10).

6.3.4. Data Coercion (Transformations)

Mathematical transformations of water quality data are usually undertaken with at least one of the following objectives in mind:

- to restore a greater degree of linearity between two or more variables,
- to stabilise the variance of some variable over time, space, or some other attribute,
- to restore a greater degree of normality in the distribution of some variable.

The identification of a suitable transformation (if it exists) is largely a trial and error process and will depend on the objectives of the exercise. Sometimes the theory or model that is being fitted to the data suggests the form of the mathematical transformation. For example, the analyst may suspect that the concentration of a nutrient is described by a power relationship of the form:

$$C = kQ^p \ ,$$

where C is concentration and Q is flow and k and p are unknown parameters. A simple logarithmic transformation yields a linear equation in $\log(C)$ and $\log(Q)$:

$$\log(C) = \log(k) + p\log(Q).$$

Thus, a log–log graph of the raw data will enable a quick assessment of the appropriateness of this model. Furthermore, estimates for $\log(k)$ and p are readily obtained as the intercept and slope respectively of the fitted regression line.

Another common goal when transforming data is to reduce the effect of a relationship between the mean and the variance. In a number of distributions (including the gamma and log-normal) the variance is related to the mean. Thus statistical tests designed to examine the equality of means will be affected by the non-constant variance. Some common transformations are provided in Table 6.5.

Table 6.5. Variance stabilising transformations (adapted from Ott 1984)

Mean (μ) ~ variance (σ^2) relationship	**Transformation**
$\sigma^2 = k\,\mu$ (Poisson data have $k = 1$)	$Y' = \sqrt{y}$ or $\sqrt{y + 0.375}$
$\sigma^2 = k\,\mu^2$	$Y' = \log(y)$ or $\log(y + 1)$
$\sigma^2 = k\,\pi(1 - \pi), \quad 0 < \pi < 1$ (Binomial data have $k = 1/n$)	$Y' = \sin^{-1}(\sqrt{y})$

It is common practice to transform data when one or more assumptions of a proposed statistical test appear to have been violated. Many analysts transform data to try and restore some semblance of normality. In many instances this is unnecessary. A number of standard statistical procedures (such as ANOVA, *t*-tests, etc.) are relatively robust in the presence of slight to moderate departures from normality.

Rather than attempting to achieve normality, the analyst should ensure that the distribution of the data has a reasonable degree of symmetry. Significant distortions in the test results are only likely to occur in cases of high skewness and/or high kurtosis. It is far more important to check that data have homogeneous variances (i.e. the variances of the variables of interest are constant over different groups, times, or space) and are independent. Data that are either spatially or temporally correlated (or both) are not amenable to the statistical test procedures described in this document.

The identification of an appropriate transformation is often a matter of trial and error. However, within the class of power transformations, Box and Cox (1964, 1982) have developed a systematic procedure. The method seeks to identify a value of the transformation parameter, λ in the expression

$$Y^{(\lambda)} = \begin{cases} \dfrac{y^{\lambda} - 1}{\lambda}, & \lambda \neq 0 \\ \ln(y), & \lambda = 0 \end{cases}$$

such that values of the transformed data ($y^{(\lambda)}$) exhibit a greater degree of normality than the original set of y values. Note: this transformation is applicable to non-negative data only. The computations associated with this process are too complex to be described here, and since the method is computationally intensive, it is best handled by computer software such as MINITAB® or S-PLUS®. See further discussion in section A5.1.2.

6.3.5. Checking Distributional Assumptions

Many statistical methods of inference rely on the assumption that the sample data have been randomly selected from a larger population of values that is normally distributed. There are good reasons why the normal distribution enjoys such a prominent role in theoretical and practical statistics. First, many naturally occurring phenomena actually exhibit normal-shaped distributions. Second, the important Central Limit Theorem in statistics assures us that even when the distribution of individual observations is non-normal, aggregate quantities (such as the arithmetic mean) will tend to have a normal distribution. Another often-used, although less convincing argument for the use of normal-based inference is that the mathematics is 'tractable' and a large number of statistical procedures have been developed around the notion of random sampling from a normally-distributed population.

The properties of the normal distribution are well known and will not be repeated here. What is important is our ability to decide if a particular set of data can be assumed to have come from some population of values whose underlying distribution is normal (or some other specified form). Before the widespread availability of computers, plotting the data on probability paper provided a check of distributional assumptions. On probability paper the axis scales are specially constructed so that

when the distributional assumption is satisfied the resulting probability plot will be linear. This is akin to plotting data on log–log scales to check the assumption of a power relationship between two variables as was done in section 6.3.4. For more information on the use of probability plots see Ott (1984).

Many environmental variables are decidedly non-normal in their distributions. River flows may be normally distributed at a single locality for a short period of time, but on broader time scales they may decrease systematically with time in the absence of rainfall (time of sampling becomes a dominating influence); and, of course, a rainfall event will have a dramatic influence on flow.

Statistical software can do the computations associated with checking distributional assumptions. Most statistical software packages have the facility for testing the assumption of normality. Other packages, such as S-PLUS®, offer more flexibility in the types of distributions that can be examined. By way of examples, consider the distribution of arsenic concentrations obtained downstream from a gold mine (Figure 6.3), and Worked Example 1, page A5-28.

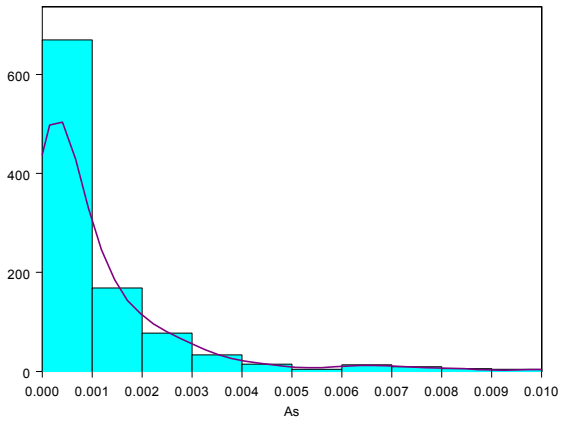


Figure 6.3. Downstream arsenic concentrations

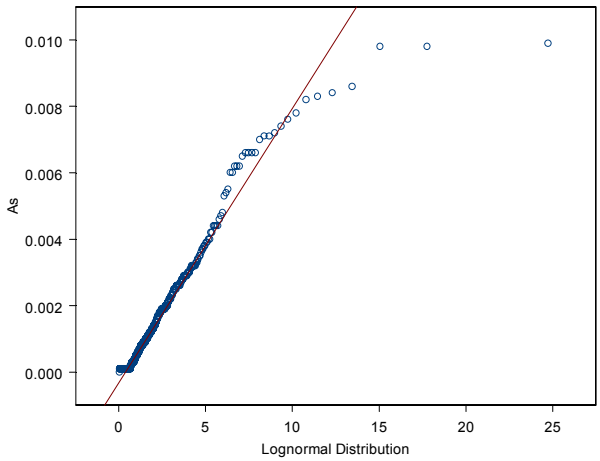


Figure 6.4. Log-normal probability plot for arsenic concentration data

Figure 6.3 reveals the non-normal shape of the data and it might be speculated that a *log-normal* distribution would be a more appropriate probability model. One way of checking the log-normal assumption would be to test the normality of the logarithms of the original arsenic data. A more direct approach is to inspect a log-normal probability plot (Figure 6.4). The tails of Figure 6.4 depart from linearity, suggesting that the data are not well described by a log-normal probability model and that a more suitable distribution may have to be sought.

6.3.6. Trend Detection

One of the principal objectives of water quality monitoring is to assess changes over time. In many instances this is driven by the need to compare water quality and a guideline value, although issues related to salinity, flow and climate variability are also important drivers.

A number of statistical methods are available to assist with trend analysis and these range from simple descriptive tools, such as time series plots, to more sophisticated modelling techniques that attempt to separate out a signal from 'noise'. These analyses have been greatly facilitated by the development of a variety of software tools. Figure 6.5a depicts arsenic time series data with the trend and 95%

prediction limits[2] overlaid. While the trend line shown is quite reasonable, care must be exercised when attempting to extrapolate beyond the range of the observed series. It is immediately apparent from Figure 6.5a that the arsenic concentrations increased quite dramatically in 1994. This could indicate altered mine practices or different ore-body characteristics and provides some explanation for the departure from linearity in the log-normal probability plot of Figure 6.4.

The trend exhibited in Figure 6.5a is revealing, but it is not the only feature of a time series with which we should be concerned. The variability of a process is equally as important as, and sometimes more important than real changes in level. Our definition of variance (Table 6.3) is based on departures from the mean (represented as the trend), so any assessment of variation should be performed on the de-trended data. The residuals (i.e. original observation minus trend) from this process can be plotted as a separate time series. This has been done for the arsenic data (Figure 6.5b).

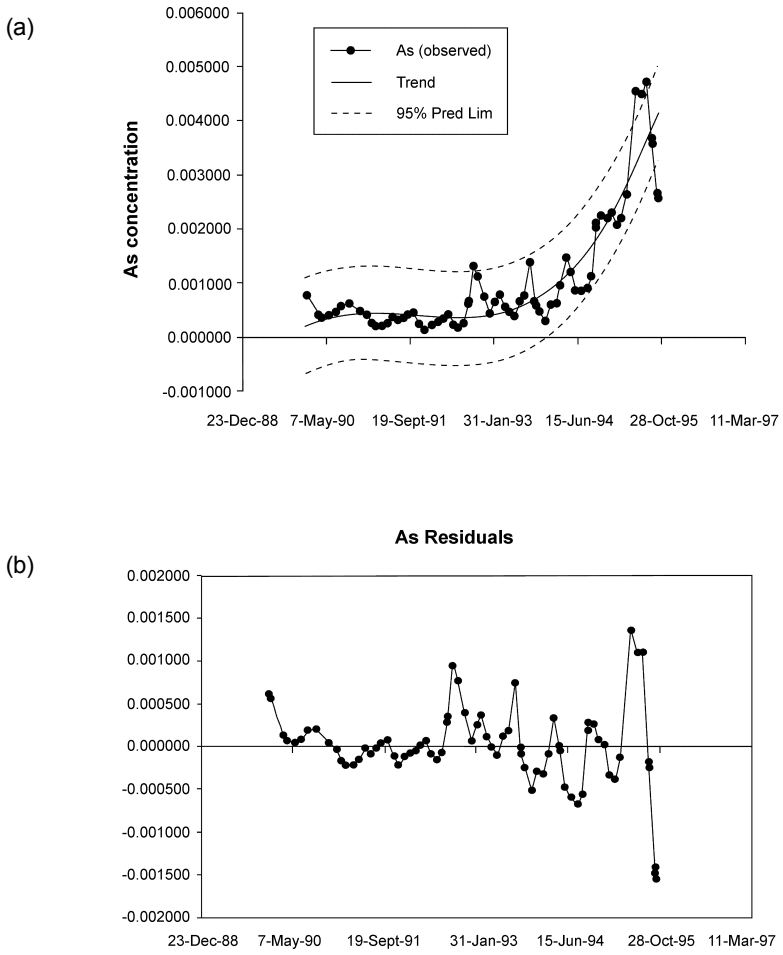


Figure 6.5. (a) Time series representation of the arsenic concentrations of Figure 6.3; (b) Time series plot of the arsenic residuals

[2] Prediction limits are used to establish bounds on an individual prediction whereas confidence limits place bounds on a mean. Since the variation associated with the prediction of a single value is greater than the variation associated with the estimation of a mean, the width of a prediction band will be greater than the width of a confidence band for some fixed certainty level.

Figure 6.5(b) shows an apparent change in the process variation beginning in 1993. Further inquiry might determine if this observation was the result of natural or human-induced effects.

6.3.7. Smoothing

Given the high variability of most natural ecosystem processes (or of indirect processes that influence them), it is not surprising that water quality data also exhibit a high degree of variation over both space and time. This high natural 'background' variation tends to mask trends in water quality parameters and reduces our ability to extract a signal from the 'noise'. Simple graphical tools such as scatterplots and time series plots can only provide a combined view of both the trend and the noise. However, so-called robust smoothers are techniques that 'let the data speak for themselves' and have been shown to be remarkably effective in teasing out a signal from very noisy data (Cleveland 1979).

Robust smoothers work by placing a 'window' over a portion of the data, computing some numerical quantity such as mean or median and then 'stepping' the window across the data and repeating the process. The collection of statistics obtained in this way can be plotted (Figure 6.6) at the mid-points of the intervals to obtain a smooth representation of the underlying process. Figure 6.6 reveals the emergence of a strong upward trend after 1994 on top of highly cyclical fluctuations. The amount of smoothing in these plots is controlled by the user who must specify a 'span' (i.e. the fraction of data to be captured in the moving window) and the number of passes over the data. Greater smoothing is achieved by increasing the span and/or the number of passes (Chambers and Hastie 1992).

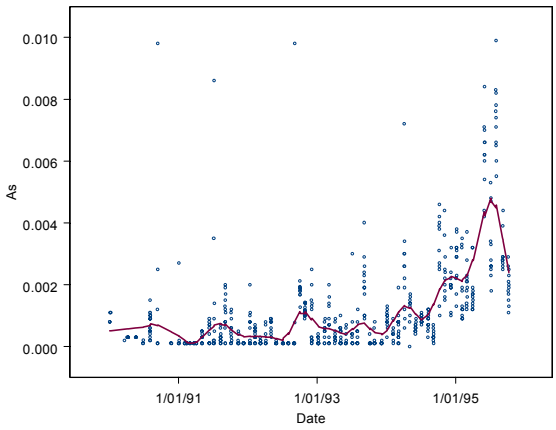


Figure 6.6. Smoothed arsenic data

6.4. Inference

The preceding sections of this chapter have focused on data analysis — that is, the process of summarising, presenting, and describing the information contained in sample information. While this is an important activity in its own right, most statistical analyses are concerned with inference — that is, methods for inferring something about some characteristic of a population of values, based on the limited information contained in a sample drawn from that population (see Table 6.9, page 6-26). In this context, statistical inference may be regarded as decision-making under uncertainty, and is thus imperfect.

6.4.1. Estimating an Unknown Water Quality Parameter

The true value of the concentration for phosphorus in a water storage is never known unless we drain the storage and measure all the phosphorus and the volume of water. Nevertheless, with an appropriate sampling regime, limits for the true value can be established. For example, the 95% confidence limits for a population mean give the range within which we can be 95% sure the true population value lies. To calculate a confidence interval for the true mean, the following formula can be used in conjunction with Table 6.6 (see also section A5.1.5.1),

$$\overline{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}},$$

where $\overline{X}$ is the sample mean, s is the sample standard deviation of n observations and $t_{\alpha/2}$ is a so-called 'critical' value from the t-distribution having degrees of freedom n–1.

Table 6.6. Critical *t*-values for selected confidence intervals and degrees of freedom (df)

df	90%	95%	99%	99.9%
1	6.314	12.706	63.657	636.619
2	2.920	4.303	9.925	31.599
3	2.353	3.182	5.841	12.924
4	2.132	2.776	4.604	8.610
5	2.015	2.571	4.032	6.869
6	1.943	2.447	3.707	5.959
7	1.895	2.365	3.499	5.408
8	1.860	2.306	3.355	5.041
9	1.833	2.262	3.250	4.781
10	1.812	2.228	3.169	4.587
11	1.796	2.201	3.106	4.437
12	1.782	2.179	3.055	4.318
13	1.771	2.160	3.012	4.221
14	1.761	2.145	2.977	4.140
15	1.753	2.131	2.947	4.073
16	1.746	2.120	2.921	4.015
17	1.740	2.110	2.898	3.965
18	1.734	2.101	2.878	3.922
19	1.729	2.093	2.861	3.883
20	1.725	2.086	2.845	3.850
21	1.721	2.080	2.831	3.819
22	1.717	2.074	2.819	3.792
23	1.714	2.069	2.807	3.768
24	1.711	2.064	2.797	3.745
25	1.708	2.060	2.787	3.725
26	1.706	2.056	2.779	3.707
27	1.703	2.052	2.771	3.690
28	1.701	2.048	2.763	3.674
29	1.699	2.045	2.756	3.659
30	1.697	2.042	2.750	3.646

For non-normal data, it is common to transform the data to yield approximate normality, then calculate confidence limits for the transformed data. It is possible to obtain approximate limits for the untransformed data by back-transformation.

Normality is not the only assumption made in using confidence limits. It is assumed that the measurements made on water samples are drawn at random from the water body in question, to protect against systematic bias. It is also assumed that the measurements are independent of each other. This might not be so if measurements are taken only from close to shore, as they will be more

similar to each other than would be measurements taken at random localities. Such measurements would be pseudoreplicates (Hurlbert 1984), and the confidence limits obtained would have an actual confidence level different to the assumed confidence level.

6.4.2. Testing Hypotheses

When a statistical hypothesis is being tested, the lack of complete information gives rise to Type I and Type II errors (Table 6.7); see section A5.1.5.2.

Statistical hypothesis tests are generally referred to as either parametric or nonparametric (i.e. 'distribution-free'). The distinction is that, in the former, the test procedure has been developed by assuming a parametric form for the underlying distribution of data values (e.g. normally distributed). The nonparametric tests relax these assumptions and are thus more robust. The price paid for this robustness is a reduction in statistical power when the assumptions of an equivalent parametric test would have been met.

It was remarked at the beginning of this chapter that the routine application of classical 'significance testing' procedures for water quality and ecological assessments is under scrutiny. Environmental researchers such as Underwood (1990, 1991, 1994), Fairweather (1991), and Green (1989, 1994) have helped raise the awareness of the need for proper statistical design and analysis in ecosystem assessment. Acronyms such as BACI, BACIP (BACI with temporal and paired spatial replication), and MBACI (BACI with full spatial and temporal replication) have helped increase this awareness. A cautionary note is nevertheless warranted. Much practical effort tends to be spent on activities that can distract the researcher from uncovering and modelling more important processes. It can be a difficult and sometimes futile exercise to attempt to identify an 'environmental control' in a landscape that has been disturbed and modified by human intervention. Similarly, the assumptions of ANOVA and related techniques have caused many analysts to think that that data must be coerced into normality at all costs.

Table 6.7. Types of error in statistical hypothesis testing

Decision	True state of nature	
	H_0 true	H_0 false
Accept H_0	✓	Type II error
Reject H_0	Type I error	✓

A balanced approach is required — one that acknowledges the rightful place of classical statistical inference yet encourages thinking and modelling outside the confines of simplistic, parametric analysis. Consistent with the recent trend away from strict 'compliance', data analysis should be focused more on discovering and understanding the spatial and temporal dynamics of an environmental impact, instead of on finding, for example, that the control and affected sites differ.

The Monitoring Guidelines cannot consider the entire range of statistical hypothesis tests in detail. Appendix 5 (e.g. sections A5.1.5.2, A5.1.6, etc.) explains the logic of significance tests and gives a more complete description of some commonly-used procedures. Two classes of statistical models that deserve greater attention in water quality studies are generalised linear models (McCullagh and Nelder 1983; Dobson 1990) and generalised additive models (Hastie and Tibshirani 1990).

Briefly, generalised linear models provide a far more flexible framework than 'conventional' methods (such as *t*-tests and ANOVA) for analysing data and making inferences. This is because of two main enhancements: (i) the error distribution is not confined to the normal probability model — i.e. log-normal, gamma, exponential, inverse Gaussian, binomial, Poisson and others are all permissible; and

(ii) generalised linear models accommodate non-linear relationships between the mean response and a set of predictor variables.

In a similar fashion, generalised additive models (or GAMs) have been devised to increase the flexibility of statistical modelling. Rather than imposing and estimating some predefined model, GAMs replace the usual linear function of an independent variable by an unspecified smooth function. In this sense, the model is nonparametric because a parametric form is not imposed on the functions — they are suggested by the data. For more discussion see section A5.1.12.

6.4.3. Comparison of Test Statistics and a Guideline or Trigger Value

This section discusses three techniques for comparing test statistics and water quality guidelines (see subsections 6.4.3.3–6.4.3.5 below). Water quality guidelines for fresh, marine, estuarine waters and groundwaters are provided in the revised Water Quality Guidelines document (ANZECC & ARMCANZ 2000), and that document also describes how the guidelines are derived and how they should be compared with test data. To give context to the discussion below, some of that background information is repeated here in sections 6.4.3.1 and 6.4.3.2.

For toxicants in water, test data are compared against default guideline values. The revised Water Quality Guidelines document (ANZECC & ARMCANZ 2000) recommends that the 95th percentile of concentration values at the test site should be less than the default guideline value for the toxicant (see Water Quality Guidelines section 7.4.4).

For biological measurement parameters and physical and chemical stressors, the revised Water Quality Guidelines advises the stakeholders for a waterbody to assess change in its quality (at its test site(s)) by comparing it with a relatively unimpacted and healthy reference waterbody or reference site(s). The reference waterbody and the test site(s) should be as similar as possible, with similar geology and climate (see Water Quality Guidelines section 3.1.4). Other matters to be considered when choosing a reference site are discussed in this Monitoring Guidelines document in section 3.2.1 and section 3.2.2, and in the Water Quality Guidelines section 3.1.4.

The manner in which test data for biological parameters are compared against reference data is discussed in the Water Quality Guidelines section 3.2.4. These issues are not discussed any further in this section.

The Water Quality Guidelines advocates that for physical and chemical (non-toxicant) parameters, the median quality values of fresh and marine waters should be lower than the 80th percentile of concentration values of a suitable reference site (above the 20th percentile for parameters such as dissolved oxygen where low values are the problem). Thus the 80th and 20th percentiles act as the trigger values (see the Water Quality Guidelines section 7.4.4).

6.4.3.1. Computation of Reference Percentiles and Their Use As Triggers

The notes in this subsection outline the derivation and use of percentiles of the reference site data for physical and chemical stressors, as described in the Water Quality Guidelines section 7.4.4. See that source for more detail.

The Water Quality Guidelines recommends that the computation of the 80th percentile at the reference site be based on the most recent 24 monthly observations there. The suggested procedure is as follows:

(i) arrange the 24 data values in ascending (i.e. lowest to highest) order; then

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

Each month, obtain a new reading at the reference (and test) sites. Append the reference site observation to the end of the original (i.e. unsorted) time sequence of reference site data, and then apply steps (i) and (ii) above to this new set of 24 data values. Note that even though only the most recent two years' data are used in the computations, no data are discarded.

A trigger for further investigation of the test waterbody will be deemed to have occurred when the *median* concentration of a particular measurement parameter in n independent samples taken at the test waterbody exceeds the 80th percentile (or is below the 20th percentile if 'less is worse') of the same measurement parameter at the reference site. A minimum of two years of consecutive monthly data at the reference site is required before a valid trigger value can be established based on that site's percentiles. If this requirement has not been satisfied, the median of the data values measured at the test site should be compared to the appropriate default guideline value identified in the Water Quality Guidelines.

The advantages of using a percentile of the reference distribution are:

- it avoids the need to specify an absolute quantity; and
- the trigger criterion is being constantly updated as the reference site is monitored, and therefore it reflects temporal trends and the effects of extraneous factors (e.g. climate variability, seasons).

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.

6.4.3.2. Number of Samples at the Test Site

The choice of number of samples (sometimes called 'sample size') to be taken 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. If the distribution of reference site values is identical to the distribution of values measured at the test site, the chance of a single observation from the test site exceeding the 80th percentile (or less than the 20th percentile) of the reference distribution is precisely 20%. Thus the Type I error (the risk of triggering a false alarm) in this case is 20%. It can be reduced by increasing n. For example, when $n = 5$ the Type I error rate is approximately 0.05. The concomitant advantage of taking more samples is the reduction in Type II error (the risk of falsely claiming a degraded water body to be 'acceptable') (see the Water Quality Guidelines section 7.4.4).

The capability of a monitoring program that compares water quality in relation to a guideline value can be described in terms of the program's power curve (Ellis 1989; Ward *et al*. 1990; Donohue 2000 unpublished). In general, a power curve is used to establish the statistical performance of a test procedure. In the present context, values of the horizontal axis represent the degree to which the reference and test sites differ with respect to a measured parameter. The vertical scale gives the corresponding power — i.e. the probability that the test procedure will correctly identify a deterioration (at the test site) of a magnitude equal to the position on the horizontal scale. Section A5.1.10 in these Monitoring Guidelines discusses the power of a sampling scheme, and lists web sites that discuss or supply software with which power and the number of samples required for its achievement can be calculated.

6.4.3.3. Comparison between Test Data and Guideline Values By Control Charts

Control charts (see also section 6.3.3) can be usefully employed for comparing data values with a guideline or trigger value, and the Water Quality Guidelines recommends that they be used.

The advantages of control charts are that:

- minimal processing of data is required;
- they are graphical — trends, periodicities and other features are easily detected;
- they have early warning capability — the need for remedial action can be seen at an early stage.

When the monthly results for a test water body are graphed in a control chart they may look like Figure 6.7(a) (compared to a trigger obtained using the 80th percentile from reference site monitoring) or Figure 6.7(b) (compared to a single published guideline value). Both these diagrams appear in the Water Quality Guidelines section 7.4.4. Confidence limits can be added to each of the

plotted points when sample means are used, but this is not straightforward when dealing with percentiles (particularly a 'rolling' percentile at a reference site).

Confidence intervals themselves provide a good way of comparing observations of a water quality variable to a guideline value; see section A5.1.5.1

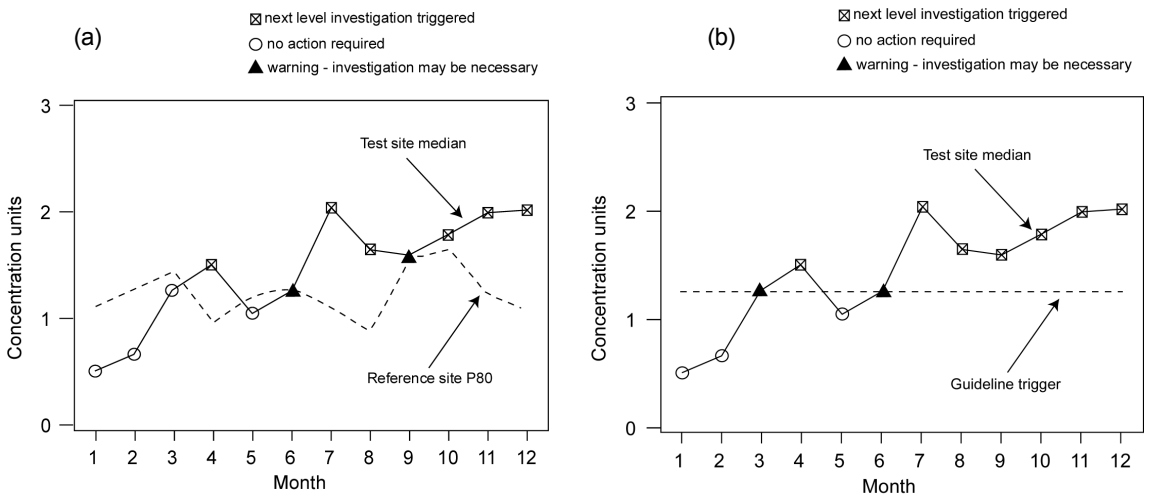


Figure 6.7. Control charts showing physical and chemical data (y-axis): (a) for test and reference sites plotted against time, and recommended actions; (b) for test site plotted against default trigger value and time, and recommended actions

6.4.3.4. Comparison between Test Data and Guideline Values by the Binomial Approach

The percentile-based method adopted by the Water Quality Guidelines for assessing water quality is a flexible nonparametric approach that is convenient and logical because inference can be made about extreme values (not just 'averages'). No assumptions have to be made about the distributional properties of the data obtained from either the test or reference sites.

A useful by-product of the percentile approach is that probabilistic assessments can be made about conformity to a guideline value without reference to the underlying statistical distribution of sample values. This requires an understanding of the binomial distribution at an introductory level.

When sampling is random and independent, variation in the number of sample excursions from a target follows the binomial distribution, regardless of the distribution of the raw data (Donohue 2000 unpublished). The binomial distribution describes the behaviour of the number of 'successes' any random process for which there are two mutually exclusive outcomes, such as heads/tails, higher than or less than. (See basic statistical texts, e.g. Walpole and Myers (1985), for details and formulae of the binomial distribution.)

Using the binomial distribution, the probability of obtaining *exactly* a certain number of samples with more (or less) than a certain level of a contaminant from a waterbody can be calculated. The binomial distribution is described completely by n, the number of random and independent samples collected, and by π, the assumed ecosystem rate of excursion. The probability of getting r samples that are worse than the trigger value ('sample excursions') from any set of n is (Ellis 1989):

$$\text{Prob.}(r) = \frac{n!}{(n-r)!r!}\pi^r(1-\pi)^{n-r} \quad (r = 0, 1, \ldots, n \;; \quad 0 < \pi < 1).$$

The binomial formula above gives probabilities for *individual* outcomes, that is probabilities associated with events of the kind 'exactly 3 out of 5 samples for which the guideline was exceeded'.

To assess the statistical significance associated with a particular outcome, a *cumulative* probability needs to be evaluated. The cumulative probability is defined as the probability of all events equal to *and more extreme than* the observed event. Thus, for example, if three out of five samples had a reading in excess of a guideline, then the appropriate cumulative probability is the probability of three or more samples (i.e. three, four or five) exceeding the guideline. This cumulative probability is known as the *p*-value in the language of hypothesis testing, and is conventionally compared to a nominal 0.05 level.

Example

Suppose the 90th percentile for cadmium in marine waters is not to exceed 5.5 µg/L. Five water samples at one location have the following Cd measurements (µg/L): 4.6, 4.4, 5.6, 4.7, 4.1.

Assuming that the 90th percentile for Cd in the marine waters is equal to 5.5 µg/L, then it can be seen that there is a 10% chance that a single reading will exceed this level. When n independent readings are taken, the probability of the number of samples r exceeding the guideline can be obtained from the binomial probability distribution. In this case (with $n = 5$ and $p = 0.1$), the expression is

$$\text{probability that } r \text{ out of 5 samples have Cd} > 5.5\ \mu\text{g/L} = \frac{5!}{r!(5-r)!}0.1^r 0.9^{5-r} \quad (r = 0,1,\ldots,5).$$

To compute a *p*-value for assessing the significance of our sample result, the probabilities of this event and those that are more extreme are computed and summed. In this case, the *p*-value is

$$\text{Prob.(1 exceedence)} + \text{Prob.(2 exceedences)} + \ldots + \text{Prob.(5 exceedences)}$$
$$= 1 - \text{Prob.(0 exceedences)} \quad = 1 - \frac{5!}{0!5!}0.1^0 0.9^{5-0} = 1 - 0.9^5 = 0.410.$$

Since this probability is much greater than the conventional 0.05 level of significance, there is insufficient evidence to assert that the guideline has been breached. In other words, it is quite probable that at least one out of five readings will exceed the 90th percentile. Note that this statement holds true irrespective of the actual numerical value of the 90th percentile or the distribution of readings. Most statistical software, including Microsoft EXCEL®, calculates binomial probabilities.

6.4.3.5. A Parametric Method for Comparing Test Data and Guideline Values

If a water quality parameter can be assumed to be normally distributed (perhaps after suitable transformation of the data), then the following procedure can be adopted (Fox 2000 unpublished).

Assume G, the guideline or trigger value, has been established so that in an 'undisturbed' system a proportion β of values will be less than G, with probability γ. Conformity to a guideline value will be demonstrated if the mean $\bar{X}$ of a sample of n readings satisfies

$$\bar{X} \le G - ks,$$

where s is the sample standard deviation and k is a factor depending on n. Values of k for selected n, β, and γ are provided by Fox (2000 unpublished) and reproduced in the table below.

Example

Referring back to the previous example, suppose it is known that the Cd levels in marine waters are normally distributed. The formula $\bar{X} \le G - ks$ uses information about the actual magnitude of the readings, not just their position relative to the 5.5 µg/L criterion. The assumption of normality is critical, as is the probability level, here called γ, because, as is shown below, conflicting results can be obtained.

For these data $\bar{X} = 4.68$ and $s = 0.563$. With $\beta = 0.90$ and $\gamma = 0.95$ and $n = 5$ we obtain $k = 3.407$ from the table below. The comparison with the guideline requires

$$\bar{X} \le 5.5 - (3.407)(0.563) = 3.582\ .$$

In this case, the sample mean of 4.68 exceeds 3.582 and, as in the example in 6.5.3.4, we are less than 95% confident that 90% of all Cd readings are below the required 5.5 µg/L.

n	$\gamma = 0.95$		$\gamma = 0.90$		$\gamma = 0.50$	
	$\beta = 0.95$	$\beta = 0.90$	$\beta = 0.95$	$\beta = 0.90$	$\beta = 0.95$	$\beta = 0.90$
2	22.261	20.583	13.090	10.253	2.339	1.784
3	7.656	6.156	5.312	4.258	1.939	1.498
4	5.144	4.162	3.957	3.188	1.830	1.419
5	4.203	3.407	3.400	2.742	1.779	1.382
6	3.708	3.006	3.092	2.494	1.751	1.361
7	3.400	2.756	2.894	2.333	1.732	1.347
8	3.188	2.582	2.754	2.219	1.719	1.337
9	3.032	2.454	2.650	2.133	1.709	1.330
10	2.911	2.355	2.568	2.066	1.702	1.324
15	2.566	2.068	2.329	1.867	1.681	1.309
20	2.396	1.926	2.208	1.765	1.671	1.301
30	2.220	1.777	2.080	1.672	1.662	1.295

However, with $\beta = 0.90$ and $\gamma = 0.50$ and n = 5 we obtain k = 1.382 from the table above. The comparison with the guideline requires:

$$\bar{X} \leq 5.5 - (1.382)(0.563) = 4.722 ,$$

which is satisfied by the observed sample mean of 4.68. This example highlights the importance of selecting the probability *a priori* — that is, *in advance of collecting the data*.

6.4.3.6. Further Discussion

For more discussion of the assessment of data against a quality threshold, see Miller and Ellis (1986), Gilbert (1987), Ellis (1989), Ward et al. (1990). Fox (2000 unpublished) and other papers are available on the CSIRO Environmental Projects Office web site, www.epo.csiro.au/library .

Donohue (2000 unpublished) is a paper that describes the binomial approach to assessing ecosystem excursions in relation to a trigger value, based on Ellis (1989). It includes a Western Australian case study in which the approach has been used. Copies can be obtained by contacting Rob Donohue, Environmental Officer, River and Estuary Investigation Section, Water and Rivers Commission, The Hyatt Centre, 3 Plain St, East Perth, WA 6004, phone (08) 9278 0586, fax (08) 9278 0532; robert.donohue@wrc.wa.gov.au.

6.5. Exploring Relationships

Relationships between pairs of water quality variables can be conveniently handled using the standard statistical tools of correlation and regression analyses.

6.5.1. Correlation Analysis

A useful adjunct to the scatterplot matrix presented in Figure 6.2 is a summary of the correlations between pairs of variables. The (Pearson) correlation coefficient is a numerical measure of the degree of linearity between two variables. Given two variables X and Y (where Y is notionally the dependent

variable and X the independent variable), the sample correlation coefficient (r) is computed using the following formula:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} .$$

Table 6.8. Broad interpretations of Pearson's correlation coefficient

Value of $\lvert r \rvert$	Interpretation
0.7 to 1.0	Strong linear association
0.5 to 0.7	Moderate linear association
0.3 to 0.5	Weak linear association
0 to 0.3	Little or no linear association

The correlation coefficient is constrained to lie in the interval $-1 \le r \le +1$. Broad interpretations of r are given in Table 6.8[3].

A correlation analysis is generally done before a more comprehensive regression analysis in which relationships between the variables are modelled and inferences are made about the likely values of true parameter values; see also section A5.1.11.

6.5.2. Regression Analysis

Usually, although not always, the objective of the regression analysis is to describe the relationship between a single dependent variable (Y) and a set of potential explanatory or independent variables (X_1, X_2, …, X_p). This is the multiple regression case. Simple linear regression refers to problems involving a single Y and a single X. Often the identification of independent and dependent variables is obvious. At other times there is no obvious labelling and an arbitrary choice can be made.

The term 'regression' covers a number of specific modelling approaches. These include:

- non-linear regression,
- multiple regression,
- stepwise regression,
- best subsets regression,
- robust regression.

An important assumption concerning the error terms of these regression models is that they are independent. Samples collected serially in time often display a degree of autocorrelation. For example, the concentration of phosphorus in storage at a particular time has a great bearing on the concentration an hour later, and probably a day later. If one measurement is well above the general trend, the other is likely to be also. Failure to ensure independence among measurements taken through time can have profound effects on the assumed Type I error rate, though the estimated parameters of the regression remain unbiased (Neter et al. 1996). One way to overcome temporal dependence is to select a sampling interval that is large enough to ensure no connection between consecutive measurements. Alternatively, various autoregressive models are available for analysing

[3] A related, and important statistic used to assess the adequacy of a fitted regression model is the coefficient of determination, R^2. In simple terms, R^2 is the proportion of total variation in some dependent variable that can be explained or accounted for by a regression model.

time series data, and the reader is referred to the text *Applied Linear Statistical Models* by Neter et al. (1996), and to Ott (1984) for an introduction.

6.5.3. Robust Regression

In terms of parameter estimation and identification of a good-fitting model**,** a significant drawback of the 'conventional' (ordinary least squares, OLS) method is its susceptibility to outliers in the data. The detection and handling of outliers was discussed in section 6.2.2 (see also section A5.1.3) and some broad recommendations were given. However, discard of observations is something that should not be undertaken lightly and needs to be fully justified. This poses a dilemma for subsequent statistical analysis given the potentially high leverage single aberrant observations may exert on the regression results. It is with these considerations in mind that alternative regression techniques have been devised which have been shown to be particularly resilient to data abnormalities. These techniques will give results which are in close agreement with classical (OLS) methods when the usual assumptions are satisfied, but differ significantly from the least-squares fit when the errors do not satisfy the normality conditions or when the data contain outliers. The statistical software package S-PLUS® has a rich suite of robust regression tools available. It includes the new Robust MM Regression, based on Rousseeuw and Yohai (1984), Yohai et al. (1991), and Yohai and Zamar (1998) as well as Least Trimmed Squares (LTS), Least Median Squares (LMS), and least absolute deviations (L1). See section A5.1.11 for more discussion.

6.5.4. High-Dimensional Data

Water quality monitoring programs often generate high-dimensional data that pose considerable challenges for the data analyst. Statistical analogues of many of the univariate techniques identified in Table 6.9(a) are available and these come under the statistical umbrella of multivariate analysis. These techniques rely heavily on a good understanding of more advanced statistical concepts and linear algebra (vectors and matrices). All the statistical software packages identified earlier in this chapter have multivariate statistical analysis capabilities. Before applying these more advanced statistical tools, the analyst should explore the data at hand and attempt to identify relationships between pairs and groups of parameters. Visualisation of data is necessarily limited to three dimensions, albeit as a two-dimensional projection. To overcome this 'curse of dimensionality', a number of statistical procedures have been devised in an attempt to reduce the number of dimensions with minimal loss of information contained in the original set of variables. One such technique that is potentially useful in water quality studies is Principal Component Analysis. The interested reader shuold consult any of the texts available on multivariate statistics.

Principal Component Analysis (PCA) constructs linear combinations of the original variables such that the resulting combinations account for the maximal amount of variation in the original data using considerably fewer constructed variables. The drawback is that the constructed variables (the 'components') generally have no real or physical meaning and have no recognisable units of measurement. However, as an exploratory analysis technique and dimension–reduction device, PCA has a potentially valuable role to play.

6.6. Changes in Space and Time

6.6.1. Time Series Analysis

Formal statistical analysis of time series data can be rather complex and requires a good degree of skill in identifying suitable models. Numerous texts have been written on time series analysis, although Cryer (1986) provides a good overview at an introductory level.

It has been previously remarked in this section that violation of the independence assumption can cause considerable distortion of the results of some standard methods of inference such as *t*-tests and ANOVA. Time series data tend to exhibit varying degrees of serial dependence. A particularly useful tool in identifying and characterising this serial dependence is the autocorrelation function or ACF. The ACF is the correlation between pairs of observations separated by a constant lag or time-step.

Further discussion of aspects of time series analysis and illustrative examples are in section A5.1.4.

6.6.2. Testing for Trend

A nonparametric test of trend that is often used in water quality studies is the seasonal Kendall test (Gilbert 1987; Helsel and Hirsch 1992).

An assumption of the trend test is that the trends are monotonic; that is, they consistently increase or decrease (Helsel and Hirsch 1992). If concentrations vary non-monotonically over the period being analysed, the results of linear tests for trend may be misleading (Robson and Neal 1996). Furthermore, it is assumed that observations are independent. When applied to data series that are not independent (that is, exhibit autocorrelation) the risk of falsely detecting a trend is increased (Ward et al. 1990; Esterby 1996). As a general rule, the level of serial correlation in a data series increases as the frequency of sampling increases. The maximum sampling frequency possible without encountering serial correlation can be thought of as the point of information saturation (Ward et al. 1990).

The seasonal Kendall test is based on the following statistic:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \text{sign}(x_j - x_k) \qquad \text{where} \qquad \text{sign}(x_j - x_k) = \begin{cases} 1 & \text{if } x_j - x_k > 0 \\ 0 & \text{if } x_j - x_k = 0 \\ -1 & \text{if } x_j - x_k < 0 \end{cases}$$

which is then compared with a tabulated critical value to assess its significance.

While the seasonal Kendall test is robust under a variety of conditions, it is nevertheless important to examine the data for outliers (see section 6.2.2) and to adjust the data for the effect of nuisance variables. For example, many water quality indicators are correlated with flow and thus the seasonal Kendall test should be applied to the residuals (i.e. flow-corrected observations). These residuals can be obtained either as data predicted, where predicted values are obtained by explicitly modelling the relationship between flow and the variable of interest (e.g. using a regression model), or by using a nonparametric smoother.

6.6.3. Multidimensional Scaling

Multidimensional scaling (MDS) is a statistical technique that attempts to reveal relationships or patterns among a large number of variables by reconstructing 'similarities' between pairs of these variables in a 'reduced' space (i.e. one of fewer dimensions); see also section A5.1.14. Biologists and ecologists have found MDS analysis particularly useful in teasing out complex space–time interactions among biological variables. Like PCA, MDS also has some limitations, including the presentation of results in an abstract space, uncertainty surrounding the appropriate dimension of the reduced space, and a multitude of similarity measures on which to base the MDS. The calculations underpinning MDS are highly complex and are iterative — that is, a terminal solution is generally found only after a number of alternatives have been explored. The lack of agreement between the similarities or distances in the final MDS representation and the original input data is measured by the so-called 'stress statistic'. The derivation of fitted distances depends also on the type of MDS that is performed. There are numerous MDS procedures, although the main dichotomy differentiates between metric MDS and non-metric MDS depending on the measurement level of the data. Metric MDS assumes the input data are quantitative (i.e. measured on an interval or ratio scale) while non-

metric MDS assumes the data are qualitative (i.e. measured on a nominal or ordinal scale)[4]. There are differing views about the extent to which MDS can be used as an inferential tool. Computationally intensive methods are available that enable the researcher to conduct formal tests of significance, although the strength of MDS for water quality studies is more likely to reside in its ability to let the analyst discern patterns and trends visually.

6.7. Interpretation

After the data analysis, the monitoring team collates the results into a concise statistical summary, and assesses these results by use of residual diagnostics (see section A5.2 Worked Examples). This is the stage at which the team interprets the information the results provide, in the context of the objectives or questions the program was originally set up to answer. Interpretations might be, for example, that the values of a contaminant exceed the guidelines for ambient water quality because of the release of effluent by a sewerage system; or that the values of important measurement parameters before and after the building of a coastal recreation development differ significantly; or that two tested factors are not significantly reducing groundwater quality.

Once the monitoring team has expressed the interpretation concisely in non-technical terms, it can decide whether or not the program objectives have been satisfied, and whether it is appropriate to report to stakeholders.

If the interpretation does not support the conceptual model or the objectives have not been met, the model needs to be refined, the study needs to be redesigned, and new or additional data need to be collected and the analysis restarted.

To assist in the evaluation of detailed monitoring programs, the study team may consider seeking an independent peer review that can assess the program design and outcomes against the monitoring program objectives.

[4] See Appendix 5, section A5.1.1, for a discussion of measurement levels.

Table 6.9(a). Summary of common statistical procedures and applications for one-variable studies

SINGLE PARAMETER INFERENCE (adapted from Kitchens 1987)				
Parameter	**Application**	**Confidence interval**	**Hypothesis**	**Test statistic**
μ (population mean)	Inference about the 'average' value of a single water quality variable. Assumes a symmetrical distribution whose tails are not excessively long	Large sample $\bar{x} \pm z_{\alpha/2} s/\sqrt{n}$	H_0: $\mu = \mu_0$	$Z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$
	Inference about the 'average' value of a single water quality variable. Assumes normally distributed data.	Small sample $\bar{x} \pm t_{\alpha/2,n-1} s/\sqrt{n}$ degrees of freedom = $n-1$		$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$
	Inference about the 'average' value of a single water quality variable. Suitable for use with symmetrical distributions with long tails.	$\bar{x}_T \pm z_{\alpha/2} s_T/\sqrt{k}$ $\bar{x}_T$ is the trimmed mean (see Table 6.2); s_T is the standard deviation of the Windsorized sample (see Table 6.3) and k is the size of the trimmed sample.		$Z = \frac{\bar{x}_T - \mu_0}{s_T/\sqrt{k}}$
$\tilde{\mu}$ (population median)	Inference about the median value of a single water quality variable. For use with skewed distributions.	Large sample ($n > 20$): 1. Arrange data in ascending order. 2. Compute $C = (n - z_{\alpha/2}\sqrt{n})/2$ and reduce to next whole number. 3. The two limits for the confidence interval are identified as the C th observations from the low end and high end of the ordered data.	H_0: $\tilde{\mu} = \tilde{\mu}_0$	$Z = \frac{2T - n}{\sqrt{n}}$ where T = no. of observations > $\tilde{\mu}_0$
		Small sample ($n < 20$): Procedure as above for large samples, except C determined from Table A5.4.		Compute p-value directly using: $p = 1 - 0.5^n \sum_{t=0}^{T-1} \binom{n}{t}$ where T = no. observations > $\tilde{\mu}_0$
π (true population proportion)	Inference about a proportion (e.g. % conformity with a target, % 'defects' in a sample). Note: p in the formulae should be a fraction between 0 and 1.	$p \pm z_{\alpha/2}\sqrt{p(1-p)/n}$	H_0: $\pi = \pi_0$	$Z = \frac{p - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}}$
σ^2 (population variance)	Inference about the variability of a single water quality variable. Assumes normally distributed data.	$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$ degrees of freedom = $n-1$	H_0: $\sigma^2 = \sigma_0^2$	$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$
ρ_0 (true correlation coefficient)	Inference about the correlation between two water quality variables. Assumes normally distributed data.		H_0: $\rho = \rho_0$	$Z = \frac{\sqrt{n-3}}{2} \ln\left[\frac{(1+r)(1-\rho_0)}{(1-r)(1+\rho_0)}\right]$

Table 6.9(b). Summary of common statistical procedures for two-variable studies

TWO PARAMETER INFERENCE (adapted from Kitchens 1987)				
Parameter	**Application**	**Confidence interval**	**Hypothesis**	**Test statistic**
μ_d (population mean difference)	Inference about the 'average' difference between pairs of a water quality variables. Assumes samples are related (e.g. 'before' and 'after') and differences are normally distributed.	$\bar{d} \pm t_{\alpha/2,n-1}\, s_d / \sqrt{n}$	$H_0: \mu_d = 0$	$t = \dfrac{\bar{d}}{s_d / \sqrt{n}}$
	Inference about the 'average' difference between pairs of water quality variables. No distributional assumptions required.		Wilcoxon signed rank test	$t = \dfrac{\bar{r}}{s_r / \sqrt{n}}$
$\mu_1 - \mu_2$	Inference concerning the difference between the means of two independent populations. Symmetrical distributions whose tails are not excessively long.	Large sample confidence interval. $(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}$	$H_0: \mu_1 = \mu_2$	$z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$
	Inference concerning the difference between the means of two independent populations. Normal distributions with equal variances.	Small sample confidence interval. $(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\ s_p \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}$ where $s_p = \sqrt{\dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}}$ and degrees of freedom $= n_1 + n_2 - 2$		$t = \dfrac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$
	Inference concerning the difference between the means of two independent populations. Symmetrical distributions with long tails.	Large sample confidence interval. $(\bar{x}_{T_1} - \bar{x}_{T_2}) \pm z_{\alpha/2} \sqrt{\dfrac{s_{T_1}^2}{k_1} + \dfrac{s_{T_2}^2}{k_2}}$ $\bar{x}_T$ is the trimmed mean (Table 6.2); s_T is the standard deviation of the Windsorized sample (Table 6.3) and k is the size of the trimmed sample.		$z = \dfrac{\bar{x}_{T_1} - \bar{x}_{T_2}}{\sqrt{\dfrac{s_{T_1}^2}{k_1} + \dfrac{s_{T_2}^2}{k_2}}}$
$\theta_1 - \theta_2$	Inference about difference between two medians. No assumptions made about shape of distributions other than they are reasonably similar.	Wilcoxon Rank Sum test (this is an ordinary *t*-test applied to the rank-transformed data, i.e. individual data values replaced by their rank, r in the respective samples). Samples of at least 10 from both populations should be available for analysis.	$H_0: \theta_1 = \theta_2$	$t = \dfrac{\bar{r}_1 - \bar{r}_2}{s_p \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ degrees of freedom $= n_1 + n_2 - 2$

Table 6.9b continued

$\pi_1 - \pi_2$	Inference about difference between two proportions (e.g. % conformity at two sites or times). Note: *p* in the formulae should be a fraction between 0 and 1.	$(p_1 - p_2) \pm z_{a/2}\ \mathrm{SE}(p_1 - p_2)$ where $\mathrm{SE}(p_1 - p_2) = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$	H_0: $p_1 = p_2$	$z = \frac{p_1 - p_2}{\sqrt{p(1-p)}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ and $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$
$\frac{\sigma_1^2}{\sigma_2^2}$	Inference about the equality of two variances. Assumes normally distributed data. *NB: The quotient* σ_1^2 / σ_2^2 *should be expressed such that the* ***larger*** *of the two (sample) variances is the numerator.*	$\frac{s_1^2}{s_2^2}\frac{1}{f_{\alpha/2,\nu_1,\nu_2}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} f_{\alpha/2,\nu_2,\nu_1}$ **NB:** note the reversal of the degrees of freedom for the critical *f* value in the left and right sides of this expression; degrees of freedom ν_1 and ν_2.	H_0: $\frac{\sigma_1^2}{\sigma_2^2} = 1$	$f = \frac{s_1^2}{s_2^2}$ critical F values are $F_{1-\alpha,\nu_1,\nu_2}$ for a 'less than' alternative hypothesis; F_{α,ν_1,ν_2} for a 'greater than' alternative hypothesis; and $F_{1-\alpha/2,\nu_1,\nu_2}$ and $F_{\alpha/2,\nu_1,\nu_2}$ for a 'not equal to' alternative hypothesis.

Chapter Seven
Reporting and Information Dissemination

7.1. Introduction

At various stages through the design and performance of the monitoring program, there will have been interaction with the end-users of the information, particularly during the setting of objectives, the detailed study design and the laboratory analyses. The monitoring team will have clearly identified the end-users' data needs and information requirements.

Once results have been obtained and interpreted, the next step is to report the findings to the people who commissioned the study. Then there can be further dissemination of the findings to other stakeholders and information users, and a system should be set up for the effective transmission of the information. Figure 7.1 is a framework for designing reporting and information dissemination systems. A checklist is presented in Table 7.1.

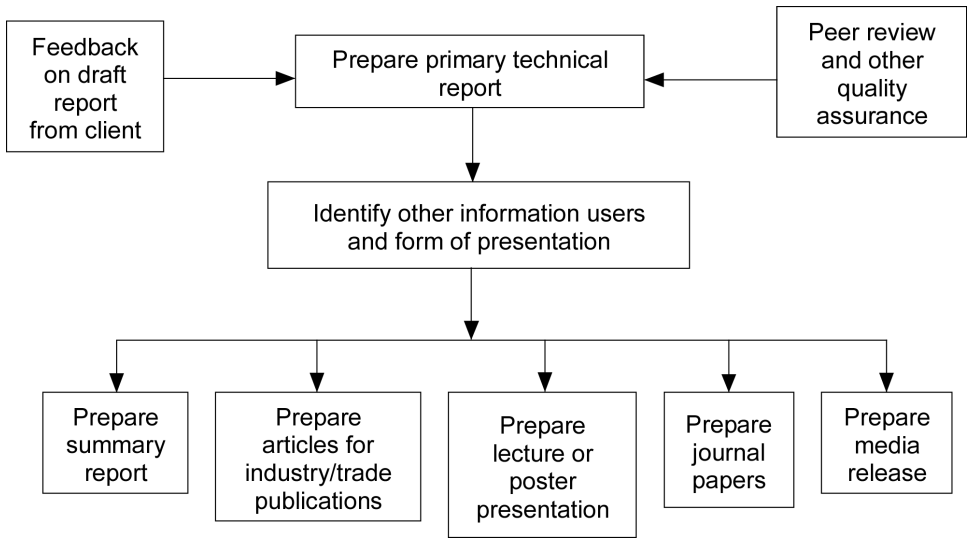


Figure 7.1. A framework for designing a reporting system

7.2. Preparation of a Primary Report

7.2.1. Reporting Schedule

The monitoring team should have negotiated a reporting schedule during the initial study design. A typical schedule might include the reporting of interim results on a monthly, bimonthly, quarterly or six-monthly basis. Such reports might be no more than data reports, with no or limited data interpretation because there will often be insufficient data for more detailed analysis. Interim reports provide a means for those conducting the program and those commissioning it to review progress and, if necessary, modify the study design.

Table 7.1. Checklist for designing a reporting system

1.	Has all the information required been obtained?
2.	Have the identities of all those who want the information been ascertained?
3.	Has the time frame in which each information user requires information been determined?
4.	Has the appropriate form of information presentation been decided?
	(a) Has the level of understanding of each user been identified?
	(b) What form of presentation will best convey the information?
	(c) Is the chosen report style consistent with the style expected by the information user?
5.	Have the available forms of information transmission been ascertained?
	(a) What form of transmission is most appropriate for each information user (written, oral, electronic)?
	(b) What procedures are in place to ensure information has reached the user?

There may be instances where the nature of the issue will make it appropriate that findings be transmitted rapidly to the end-users of the data, e.g. where there is a public health concern such as with drinking or bathing water. Rapid reporting will allow the nature of any problems to be identified early and may ensure the rapid initiation of remediation activities. In such instances, a rapid interim communication could be made, with a detailed report to follow. It is important that the pressure for results does not compromise essential quality assurance and checks of the analytical data and their analysis.

7.2.2. Report Format

At the completion of the study, or usually on an annual basis for on-going monitoring programs, a primary report is produced, in a form agreed to by all parties.

There are traditionally accepted styles for reporting the results of monitoring programs. The reports have a number of essential elements which make the information of value to all users:

- an Executive Summary that expresses the technical findings in relation to the objectives, succinctly and in words that are understandable by managers unfamiliar with technical detail;
- an introduction, outlining previous studies in the area or related studies, and delineating the study objectives;
- experimental detail, describing the study location and study design, including descriptions of methods of sampling and analysis;
- results — descriptive and detailed presentation of results, sometimes in combination with the Discussion section;
- discussion of the results including data interpretation and implications for management;
- conclusions drawn from the results;
- recommendations for future work;
- reference details for literature cited in the report;
- appendices, providing laboratory reports, data tables or other information that is too detailed or distracting to be included in the main body of the report.

Data summaries and graphical presentations of results significantly enhance the readability and utility of reports; they are excellent ways of presenting the data.

The primary report contains full and complete details of all aspects of the study. It includes detail of the sampling locations so they can be unambiguously identified, e.g. GPS directions. It includes information that identifies biota by size (length or weight) and sex, where this is appropriate. For sediment data, it reports grain size (at least the percentage of clay/silt <63 µm) and, where appropriate, organic carbon content. These finer points of detail make it possible to compare the data with those from other studies; without the detail, the comparisons will have little significance.

The primary report is the reference or source document for subsequent publications. It is important that the clients receive a draft version of the document to ensure that the final product will meet their expectations and cover all the issues that they had hoped to see addressed. This is the first stage of the review process.

Most organisations also have an internal review process that may occur before or after the client review, to ensure the quality of the output from the perspective of the organisation.

When a report is very large, a shorter summary report may be useful, to encapsulate the technical findings of each component of the study in an abbreviated form.

So that the client and other stakeholders can have confidence in the output, it is desirable to organise external peer review of the report. Such a review should address data quality, data interpretation, and whether the findings are scientifically acceptable and consistent with the monitoring program objectives.

The review may involve a technical review panel associated with the project, or a single expert. In all cases it is important that the reviewer be independent from those associated with the study.

7.3. Identifying Users and their Information Requirements

The dissemination of information to other users is not necessarily planned in advance. It can be organised after the delivery of the primary report, if necessary. The primary report would normally be made available to the clients and to a distribution list that they would nominate. Further, depending upon the commercial sensitivity of the information, the report could be made publicly available, as a result of national, regional or organisational listings of publications or other bibliographic services. Being a technical document the primary report might be too complex for other than a technical audience, and less technical versions might be needed, to describe the study outcomes, or to seek other means of disseminating information.

There is a view that where data are collected with public funds, or required as part of a licence, they should be publicly available. Some jurisdictions are now placing such data on web sites to make them widely available with minimal data extraction costs. For example, the Victorian State Water Resources Data Warehouse is available at www.vicwaterdata.net/. It is intended to improve community access to information about Victoria's water resources, particularly the water quality and stream-flow monitoring programs, and ultimately to provide information about marine, biological, groundwater and community monitoring.

A broad range of stakeholders and others will use the information that the monitoring program provides. They may have been involved in the study design, and could include:

- the resource manager (state, local government, catchment manager), concerned with the health of the catchment and management of the impacts of catchment activities or remedial works to ensure that desired and expected beneficial effects of investment are achieved;
- an environmental agency, needing to assess data in terms of trends, comparisons with other waterbodies, reporting for State of the Waters, or State of the Environment reports;
- individual water users, who are usually concerned when the values of particular measurements fail to remain within safe operating limits, and who need advice when concentrations depart from this range;
- industries that use the water or discharge into it; they will need to consider the significance of the results to their operations;
- community groups, similar to the water user but generally with broader interest in comparing their waterbody with a standard, and concern about catchment management, integrated land and water use, etc.;
- the general public (not necessarily local), who receive general media reports;
- the scientific community, concerned as part of particular environmental research.

The scope and format of reporting to these users will differ in complexity and technical content, depending on the users and their concerns. Each group will have different levels of understanding and technical skills. It is important that the level of understanding of each user group be identified, and that the information be conveyed to them in the most appropriate form.

7.4. Information Transmission

Methods for information reporting and dissemination can be categorised as follows:

- publications — these include technical reports and CD-ROMs, papers for scientific journals or books, articles in industry, trade or society journals (water industry, scientific instrumentation, trade, etc.), newsletters;
- industry and professional association conferences, seminars and workshops, community group presentations and training, open-day activities and demonstrations — platform presentation with visual aids (overhead transparencies, slides, PowerPoint presentations);
- Internet web pages;
- film and video presentations;
- media releases and media articles.

7.4.1. Publications

The primary report and scientific publications contain the detailed data interpretation and analysis in a well-established format. They serve a purely technical audience. There is general concern that technical reports are not made available to the wider scientific community, largely because of limited circulation, severely limiting the sharing of such findings. It is important to ensure that such material is abstracted by abstracting services such as Streamline.

CD-ROMs provide an excellent way of storing complex data sets in a form that enables wider data use. Currently the availability of CD-ROMs is similar to that of technical reports.

The scope for publishing the findings of monitoring programs in scientific journals is often limited, because the findings are typically only of local interest. In most cases, only those studies that use novel methods, or add to the understanding of particular environmental processes and systems, will be accepted for publication.

Written presentations to national or regional conferences or workshops are useful vehicles for publicising monitoring studies, as are more general articles that can be published in trade journals or journals of relevant industry groups or scientific societies.

Many organisations publish newsletters that include short reports of their activities, and these can be useful for publicising a monitoring program undertaken by the organisation or funded by it. These newsletter reports are excellent for raising awareness about the monitoring program, and can form the basis of a press release to trade journals and the media generally. Other printed material that publicises findings falls into this category also.

7.4.2. Meeting Presentations

Information about the monitoring program can be presented to a conference or seminar. The standard of both platform and poster presentations at conferences has increased dramatically in recent years, because improved software packages allow slides and graphic design to be prepared on a personal computer. Guidance on the latest packages is freely available, and most conference organisers will advise on acceptable standards of presentation.

The information can be presented to community groups at their regular meetings or training sessions, either in-house or externally. Internal activities can also include open days and associated

demonstrations at which invited visitors or the general public can gain information about monitoring activities.

7.4.3. Internet Web Pages

The use of Internet web pages is a powerful means of making data available to a very wide user audience. Many agencies and industries are now making reports and monitoring data available this way. However, there is concern about how the data might be used. It is pointless to simply list the monitoring data without also providing associated professional interpretation, because the data could be given potentially misleading interpretation by non-professionals. The same will apply to reporting in the media.

7.4.4. Film and Video Presentations

Film or, more commonly, video presentations are complex and expensive ways of reporting on a monitoring study. They are usually unable to convey sufficient technical detail, and are more useful for publicity than for communicating the detail of scientific findings.

7.4.5. Media Reporting

Media reports are important for disseminating general information about a monitoring program, and can come about in a structured (controlled) or unstructured way. In a structured way, the information is made available via a media release that has been produced on behalf of the organisation undertaking the study and/or the client of the study. This release is usually the responsibility of an officer with training in communicating information. Media releases will name a contact officer for further information, generally the communications officer and possibly a senior person responsible for the report or program. It is highly desirable that personnel listed on press releases or otherwise involved in contacts with the media should have training in such activities.

Irresponsible, unstructured reporting of environmental findings can lead to undue public anxiety and the wasted effort of agencies in response to political pressures, readdressing issues that have already been covered as part of the investigation but incorrectly reported. Personnel being interviewed by the print media should ask to view a transcript of the article before it is published.

Appendix 1
National Water Quality Management Strategy

The Australian and New Zealand Environment and Conservation Council (ANZECC) and the Agriculture and Resource Management Council of Australia and New Zealand (ARMCANZ) are working together to develop the National Water Quality Management Strategy (NWQMS). The National Health and Medical Research Council (NHMRC) is involved in aspects of the NWQMS which affect public health.

The NWQMS has three major elements: policies, process and guidelines.

Policies

The main policy objective of the NWQMS is set out in NWQMS Paper No. 2, *Policies and Principles — A Reference Document* (ANZECC & ARMCANZ 1994) and is

> to achieve sustainable use of the nation's water resources by protecting and enhancing their quality while maintaining economic and social development.

This objective is being pursued through a strategy based on high-status national guidelines with local implementation.

Policies and Principles — A Reference Document emphasises the importance of:

- ecologically sustainable development;
- integrated (or total) catchment management;
- best management practices, including the use of acceptable modern technology and waste minimisation and utilisation; and
- the role of economic measures, including 'user-pays' and 'polluter-pays' approaches.

Process

The process for water quality management starts with the community working in concert with government to develop a management plan for each catchment, aquifer, estuary, coastal water or other waterbody. The plan should take account of all existing and proposed activities and developments; it should contain feasible management options that aim to achieve the environmental values that have been agreed for that waterbody. The process is outlined in NWQMS Paper No. 3, *Implementation Guidelines* (ANZECC & ARMCANZ 1998) and is schematically represented in Figure A1.1. The NWQMS envisages use of both regulatory and market-based approaches.

Management of water resources is mainly a state and territory responsibility, but the NWQMS will be implemented in the context of:

- the NWQMS guidelines;
- state and territory water policies;
- community preferences on the use and values of local waters;
- the current water quality of local waters; and
- the economic and social impacts of maintaining current water quality or of meeting new local water quality goals.

Implementation of the NWQMS should include:

- catchment, groundwater and coastal water quality management plans;
- an appropriate level of water and sewerage services provided by water authorities; and
- further development of regulatory and market frameworks.

Community views form a crucial part of the NWQMS and public comment is sought during both the development and the implementation of the strategy.

Environmental values and water quality (NWQMS) guidelines are described in the *Australian and New Zealand Guidelines for Fresh and Marine Water Quality* (ANZECC & ARMCANZ 2000).

National Guidelines

The national guidelines are technical papers providing guidance on many aspects of the water cycle including ambient and drinking water quality, monitoring, groundwater, rural land and water, urban stormwater, sewerage systems and effluent management for specific industries. The full list of NWQMS documents, with their current status (19 completed so far out of 21), is in Table A1.1. The list, together with other information, is also on the NWQMS web site at http://www.affa.gov.au/nwqms.

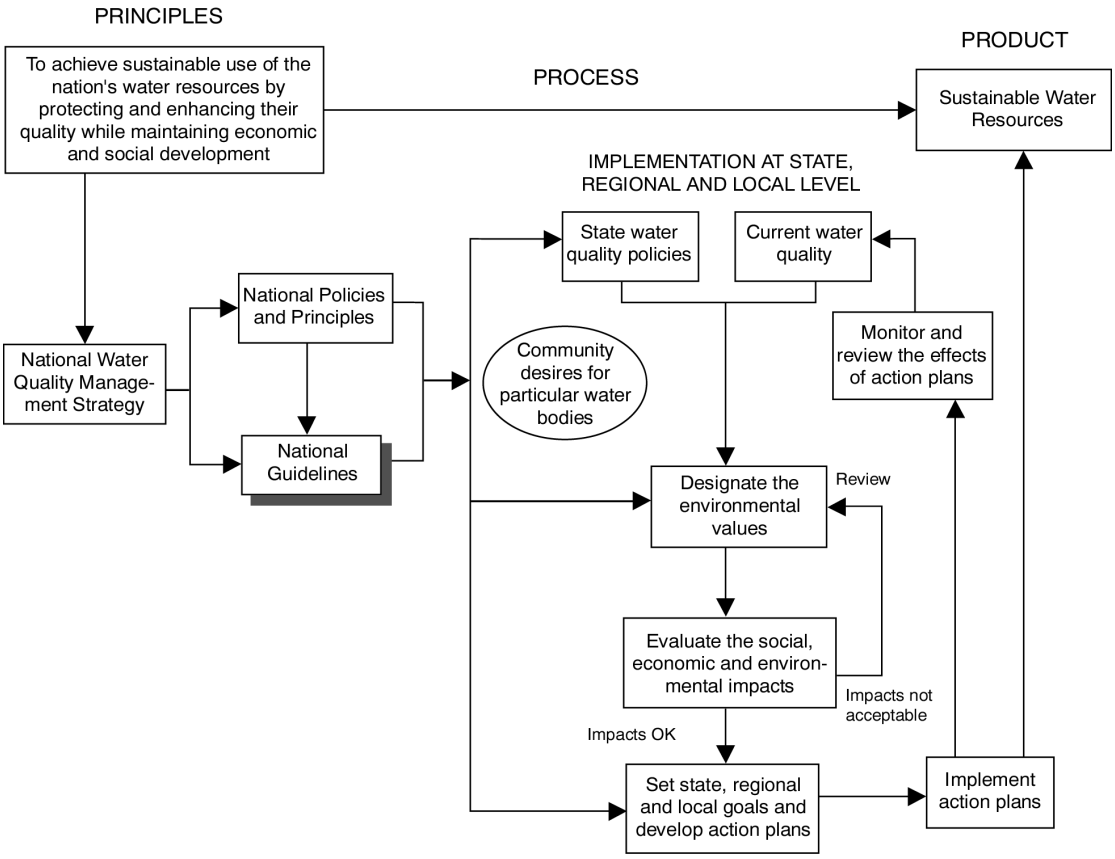


Figure A1.1. National Water Quality Management Strategy

Table A1.1. The technical papers of the National Water Quality Management Strategy, by category

Policies and Process for Water Quality Management

Paper no. 1. *Water Quality Management — An Outline of the Policies*

Paper no. 2. *Policies and Principles — A Reference Document*

Paper no. 3. *Implementation Guidelines*

Water Quality Benchmarks

Paper no. 4. *Australian and New Zealand Guidelines for Fresh and Marine Water Quality*

Paper no. 4a. *An Introduction to the Australian and New Zealand Guidelines for Fresh and Marine Water Quality*[S]

Paper no. 5. *Australian Drinking Water Guidelines — Summary*

Paper no. 6. *Australian Drinking Water Guidelines*

Paper no. 7. *Australian Guidelines for Water Quality Monitoring and Reporting*

Paper no. 7a. *Australian Guidelines for Water Quality Monitoring and Reporting — Summary*[S]

Groundwater Management

Paper no. 8. *Guidelines for Groundwater Protection*

Guidelines for Diffuse and Point Sources*

Paper no. 9. *Rural Land Uses and Water Quality — A Community Resource Document*

Paper no. 10. *Guidelines for Urban Stormwater Management*

Paper no. 11. *Guidelines for Sewerage Systems — Effluent Management*

Paper no. 12. *Guidelines for Sewerage Systems — Acceptance of Trade Waste (Industrial Waste)*

Paper no. 13. *Guidelines for Sewerage Systems — Sludge (Biosolids) Management*[#]

Paper no. 14. *Guidelines for Sewerage Systems — Use of Reclaimed Water*

Paper no. 15. *Guidelines for Sewerage Systems — Sewerage System Overflows*[#]

Paper no. 16a. *Effluent Management Guidelines for Dairy Sheds*

Paper no. 16b. *Effluent Management Guidelines for Dairy Processing Plants*

Paper no. 17. *Effluent Management Guidelines for Intensive Piggeries*

Paper no. 18. *Effluent Management Guidelines for Aqueous Wool Scouring and Carbonising*

Paper no. 19. *Effluent Management Guidelines for Tanning and Related Industries in Australia*

Paper no. 20. *Effluent Management Guidelines for Australian Wineries and Distilleries*

*The guidelines for diffuse and point sources are national guidelines that aim to ensure high levels of environmental protection that are broadly consistent across Australia.

[#] Not yet released in final form

[S] This document is available with its main document, but not as a separate item.

Appendix 2

The Council of Australian Governments' Water Reform Framework

In 1994, the Council of Australian Governments (COAG) met to discuss water resource policy, including the establishment of a Water Reform Framework, as outlined in *Managing Australia's Inland Waters: Roles for Science and Technology* (a report to the Prime Minister's Science and Technology Council, Commonwealth of Australia 1996, Appendix 3, pages 136–142). The Council comprises the Prime Minister, Premiers and Chief Ministers and the President of the Local Government Association. The detailed decisions of the Council were as below.

Water Resource Policy

In relation to water resource policy, the Council agreed:

1. that action needs to be taken to arrest widespread natural resource degradation in all jurisdictions occasioned, in part, by water use and that a package of measures is required to address the economic, environmental and social implications of future water reform;

2. to implement a strategic framework to achieve an efficient and sustainable water industry comprising the elements set out in (3) through (8) below;

3. in relation to pricing:

(a) in general —

(i) to the adoption of pricing regimes based on the principles of consumption-based pricing, full-cost recovery and, desirably, the removal of cross-subsidies which are not consistent with efficient and effective service, use and provision; where cross-subsidies continue to exist, they be made transparent,

— Queensland, South Australia and Tasmania endorsed these pricing principles but have concerns on the detail of the recommendations,

(ii) that where service deliverers are required to provide water services to classes of customer at less than full cost, the cost of this be fully disclosed and ideally be paid to the service deliverer as a community service obligation;

(b) for urban water services —

(i) to the adoption by no later than 1998 of charging arrangements for water services comprising an access or connection component together with an additional component or components to reflect usage where this is cost-effective,

(ii) that in order to assist jurisdictions to adopt the aforementioned pricing arrangements, an expert group, on which all jurisdictions are to be represented, report to COAG at its first meeting in 1995 on asset valuation methods and cost-recovery definitions, and

(iii) that supplying organisations, where they are publicly owned, aim to earn a real rate of return on the written-down replacement cost of their assets, commensurate with the equity arrangements of their public ownership;

(c) for metropolitan bulk-water suppliers —

(i) to charging on a volumetric basis to recover all costs and earn a positive real rate of return on the written-down replacement cost of their assets;

(d) for rural water supply —

(i) that where charges do not currently fully cover the costs of supplying water to users, that charges and costs be progressively reviewed so that no later than 2001 they comply with the principle of full-cost recovery with any subsidies made transparent consistent with 3(a)(ii) above,

(ii) to achieve positive real rates of return on the written-down replacement costs of assets in rural water supply by 2001, wherever practicable, that future investment in new schemes or extensions to existing schemes be undertaken only after appraisal indicates it is economically viable and ecologically sustainable,

(iii) where trading in water could occur across State borders, that pricing and asset valuation arrangements be consistent,

(iv) where it is not currently the case, to the setting aside of funds for future asset refurbishment and/or upgrading of government-supplied water infrastructure, and

(v) in the case of the Murray-Darling Basin Commission, to the Murray-Darling Basin Ministerial Council putting in place arrangements so that, out of charges for water, funds for the future maintenance, refurbishment and/or upgrading of the headworks and other structures under the Commission's control be provided;

(e) for groundwater —

(i) that management arrangements relating to groundwater be considered by the Agriculture and Resource Management Council of Australia and New Zealand (ARMCANZ) by early 1995 and advice from such considerations be provided to individual jurisdictions and the report be provided to COAG.

4. In relation to water resource policy, the Council agreed, in relation to water allocation or entitlements:

(a) that State Government members of the Council would implement comprehensive systems of water allocations or entitlements backed by separation of water property rights from land title and clear specification of entitlements in terms of ownership, volume, reliability, transferability and, if appropriate, quality,

(b) where they have not already done so, States would give priority to formally determining allocations or entitlements to water, including allocations for the environment as a legitimate user of water,

(c) in allocating water to the environment, member governments would have regard to the work undertaken by ARMCANZ and the Australian and New Zealand Environment and Conservation Council (ANZECC) in this area,

(d) that the environmental requirements, wherever possible, will be determined on the best scientific information available and have regard to the inter-temporal and inter-spatial water needs required to maintain the health and viability of river systems and groundwater basins. In cases where river systems have been over-allocated, or are deemed to be stressed, arrangements will be instituted and substantial progress made by 1998 to provide a better balance in water resource use including appropriate allocations to the environment in order to enhance/restore the health of river systems,

(e) in undertaking this work, jurisdictions would consider establishing environmental contingency allocations which provide for a review of the allocations five years after they have been determined, and

(f) where significant future irrigation activity or dam construction is contemplated, appropriate assessments would be undertaken to, inter alia, allow natural resource managers to satisfy

themselves that the environmental requirements of the river systems would be adequately met before any harvesting of the water resource occurs.

5. In relation to water resource policy, the Council agreed, in relation to trading in water allocation or entitlements:

(a) that water be used to maximise its contribution to national income and welfare, within the social, physical and ecological constraints of catchments,

(b) where it is not already the case, that trading arrangements in water allocations or entitlements be instituted once the entitlement arrangements have been settled. This should occur no later than 1998,

(c) where cross-border trading is possible, that the trading arrangements be consistent and facilitate cross-border sales where this is socially, physically and ecologically sustainable, and

(d) that individual jurisdictions would develop, where they do not already exist, the necessary institutional arrangements, from a natural resource management perspective, to facilitate trade in water, with the proviso that in the Murray-Darling Basin the Murray-Darling Basin Commission be satisfied as to the sustainability of proposed trading transactions.

6. In relation to water resource policy, the Council agreed, in relation to institutional reform:

(a) that where they have not already done so, governments would develop administrative arrangements and decision-making processes to ensure an integrated approach to natural resource management,

(b) to the adoption, where this is not already practised, of an integrated catchment management approach to water resource management, and to set in place arrangements to consult with the representatives of local government and the wider community in individual catchments,

(c) to the principle that, as far as possible, the roles of water resource management, standard setting and regulatory enforcement and service provision be separated institutionally,

(d) that this occur, where appropriate, as soon as practicable, but certainly no later than 1998,

(e) to the need for water services to be delivered as efficiently as possible and that ARMCANZ, in conjunction with the Steering Committee on National Performance Monitoring of Government Trading Enterprises, further develop its comparisons of inter-agency performance, with service providers seeking to achieve international best practice,

(f) that the arrangements in respect of service delivery organisations in metropolitan areas in particular should have a commercial focus, and, whether achieved by contracting-out, corporatised entities or privatised bodies, this be a matter for each jurisdiction to determine in the light of its own circumstances, and

(g) to the principle that constituents be given a greater degree of responsibility in the management of irrigation areas, for example, through operational responsibility being devolved to local bodies, subject to appropriate regulatory frameworks being established.

7. In relation to water resource policy, the Council agreed, in relation to consultation and public education:

(a) to the principle of public consultation by government agencies and service deliverers where change and/or new initiatives are contemplated involving water resources,

(b) that where public consultation processes are not already in train in relation to recommendations (3)(b), (3)(d), (4) and (5) in particular, such processes will be embarked upon,

(c) that jurisdictions individually and jointly develop public education programs in relation to water use and the need for, and benefits from, reform,

(d) that responsible water agencies work with education authorities to develop a more extensive range of resource materials on water resources for use in schools, and

(e) that water agencies should develop, individually and jointly, public education programs illustrating the cause and effect relationship between infrastructure performance, standards of

service, and related costs, with a view to promoting levels of service that represent the best value for money to the community.

8. In relation to water resource policy, the Council agreed, in relation to the environment:

(a) that ARMCANZ, ANZECC and the Ministerial Council for Planning, Housing and Local Government examine the management and ramifications of making greater use of waste water in urban areas and strategies for handling stormwater, including its use, and report to the first Council of Australian Governments meeting in 1995 on progress,

(b) to support ARMCANZ and ANZECC in their development of the National Water Quality Management Strategy, through the adoption of a package of market-based and regulatory measures, including the establishment of appropriate water quality monitoring and catchment management policies and community consultation and awareness,

(c) to support consideration being given to establishment of landcare practices that protect areas of river which have a high environmental value or are sensitive for other reasons, and

(d) to request ARMCANZ and ANZECC, in their development of the National Water Quality Management Strategy, to undertake an early review of current approaches to town waste water and sewage disposal to sensitive environments, noting that action is under way to reduce accessions to water courses from key centres on the Darling River system (it was noted that the National Water Quality Management Strategy is yet to be finalised and endorsed by governments).

9. In relation to water resource policy, the Council agreed, in relation to water and related research:

(a) to give higher priority to the research necessary to progress implementation of the strategic framework, including consistent methodologies for determining environmental flow requirements, and

(b) to greater coordination and liaison between research agencies to more effectively utilise the expertise of bodies such as the Land and Water Resources Research and Development Corporation, the Murray-Darling Basin Commission and other State and Commonwealth organisations.

10. In relation to water resource policy, the Council agreed, in relation to taxation:

(a) that a sub-committee of Commonwealth and State officials, established by the Working Group on Micro-economic Reform, meet to discuss taxation issues of relevance to the water industry with a view to reporting, through the Working Group, to the Council within 12 months,

(b) to support water-related taxation issues being examined in the proposed Industry Commission Inquiry into Private Sector Infrastructure Funding, and

(c) to accept any future consideration of tax compensation payments involving the water industry being dealt with through the Commonwealth–State Working Group established at the July 1993 financial Premiers' Conference.

11. In relation to water resource policy, the Council agreed, in relation to recommendations (3) through (8):

(a) that the Working Group on Water Resource Policy would coordinate a report to the Council for its first meeting in 1995 on progress achieved in implementing this framework including reductions in cross-subsidies, movement towards full-cost recovery pricing in urban and rural areas and the establishment of transferable water entitlements, and

(b) that as part of the monitoring and review process, ARMCANZ, ANZECC and, where appropriate, the Murray-Darling Basin Ministerial Council and the Ministerial Council for Planning, Housing and Local Government, would report annually over the succeeding four years, and again at its first meeting in 2001, to the Council of Australian Governments on progress in implementing the various initiatives and reforms covered in this strategic framework.

Appendix 3
Current Monitoring Approaches

A3.1. Background

Water quality monitoring is undertaken by Commonwealth, state, territory and local governments, universities, research organisations, private companies, schools and community organisations.

The Environment Protection Group of Environment Australia commissioned Aquatech Pty Ltd to investigate water quality monitoring in Australia (DEST 1995). This work has since been updated in a new survey (EA/NLWRA 2000). The consultant noted that water quality monitoring is relatively well organised, although there is scope for improvement in a number of areas, such as better coordination and design of monitoring programs. In 1993–94, about 1500 distinct water quality monitoring programs were operated by 214 organisations (EA/NLWRA 2000). Mining companies did a significant amount of monitoring in Western Australia and the Northern Territory. In 1998–99, about 1000 programs were in progress, operated by 118 organisations, but it is probable that former small programs had been rationalised into larger programs, causing the apparent reduction.

The numbers of water quality monitoring programs for various categories of water are given in Table A3.1. Of these programs, 72% were conducted by state, territory or local governments in 1993–94 compared to 90% in 1998–99. The total cost for all organisations in 1993–94 was estimated by Aquatech to be around $60 million, while in 1998–99 it was $112 million (EA/NLWRA 2000).

Table A3.1. Numbers of water quality monitoring programs across Australia in 1993–94 and in 1998–99 conducted by governments, universities, research organisations, private companies and community organisations

Category	Number in 1993–94	Number in 1998–99
Total	1489	999
Some of the types of waterbody monitored		
Drinking water supply reservoirs	174	n.a.
Reservoirs and lakes	n.a.	113
Lakes and dams	154	n.a.
Rivers and creeks	210	291
General riverine environments	89	n.a.
Urban stormwater	43	14
Groundwaters	112	69
Estuaries	48	83
Coastal waters	52	46
Industrial water supply and process effluent	321	n.a.
Industrial effluent	n.a.	177
Agricultural runoff	17	6

The community is playing an increasing role with programs such as Waterwatch, and its contribution is vital. The majority of waters tested tend to be industrial water supplies and effluent, followed by drinking water reticulated systems, rivers and streams, drinking water reservoirs, dams, groundwaters, riverine environs and coastal waters; agricultural runoff waters are the least tested. Currently about 20% of monitoring programs monitor biological parameters. In the current programs, attached algae, macrophytes, invertebrates, fish, birds and animals are tested. About 23% of monitoring programs include water flow measurements. These data can assist in determining nutrient loads and trends.

A3.2. Stakeholders Involved in Water Quality Monitoring in Australia

The roles of several stakeholders in the water quality monitoring industry are outlined below. Stakeholders may both provide and use information from water quality monitoring.

A3.2.1. Commonwealth Government

The Commonwealth Government has an interest in water quality issues because of their potential significance to the national economy. The Government is concerned that the community has access to clean water, and it has international responsibilities for the maintenance of environmental resources. It also recognises that national involvement and leadership help ensure that the monitoring actions taken are coordinated and in the national interest.

The Commonwealth provides funding for research and ongoing water quality monitoring programs by the states and the community, and for educational material: for example, it runs the national State of the Environment Reporting System and the Waterwatch national office as well as water research institutions.

The Commonwealth has direct responsibility for meteorology and the Exclusive Economic Zone. It cooperates with states and territories on water issues through Councils such as ANZECC, ARMCANZ and NHMRC. The Commonwealth's role in water quality monitoring is maintained through its involvement in the National Water Quality Management Strategy (NWQMS), through the promotion of consistent approaches to monitoring across Australia, and information dissemination.

Significant research programs on water issues are conducted by CSIRO, the universities and several Cooperative Research Centres (CRCs). The main water research brokering and funding body is the Land and Water Resources Research and Development Corporation (LWRRDC), which covers the rural environment. Urban water research issues are covered by the Urban Water Research Association of Australia (UWRAA). (The UWRAA is not a Commonwealth body, but is a Division of the Water Services Association of Australia.)

A3.2.2. State and Territory Governments

The management of Australia's water resources is the responsibility of individual state and territory governments. State and territory government water resource agencies are responsible for the monitoring of ambient water quality in streams and storages. However, the setting of water quality standards and guidelines, ensuring that specific monitoring of waste discharges is carried out, is the responsibility of environmental protection agencies which also run monitoring programs. To meet these obligations, in part, state and territory agencies monitor ambient water quality throughout their states as well as carrying out specific monitoring in government-owned irrigation areas and projects. Governments generally have an interest in the monitoring of water quality, not only for its own intrinsic value, but also because water quality has the potential to be an overall indicator of the condition of catchments and the effectiveness of integrated natural resource management and catchment management.

The day-to-day management of pollution control and water quality monitoring is the responsibility of state and territory governments in accordance with their responsibilities under the Australian Constitution. State, territory and also local governments conduct the majority of water quality monitoring programs.

Water administration in major urban areas is done by water management authorities, such as Sydney Water and Melbourne Water, which have responsibility for treatment, supply and sewerage.

The environment protection agencies regulate industry, manage legislative responsibilities and determine conditions for industry compliance, policing and monitoring. Also, there is a range of other state and territory departments, such as public works and services, health, agriculture, fisheries, mines, national parks, conservation, and planning, which are involved to some degree in environmental management leading to water quality monitoring.

A3.2.3. Local Governments

Local governments are increasingly involved in issues relating to environmental management within their jurisdictions, particularly in the planning context. In addition, in some of the eastern states of Australia, local councils are responsible for the provision of water and sewerage and stormwater disposal. Adequate monitoring and reporting is essential for ensuring that these responsibilities are discharged in a safe and environmentally sensitive manner.

A3.2.4. Universities and Research Organisations

Contributions from universities and other research organisations include the development of water monitoring techniques, and the uptake of new technology and research into a wide range of related issues. They generate a significant amount of raw data on a range of water quality issues and techniques. These bodies interface with industry and governments through the CRCs, and assist in information dissemination and fund a variety of water-related research and training programs.

A3.2.5. Industry

Operators in various industries are often required to collect data regarding waste streams and ambient water conditions on an ongoing basis. In addition, operators in industries that pose a potential threat to surface or groundwater are often required to monitor, to ensure the integrity of their operations. Increasingly, industry is encouraged to be self-regulating, backed up by regular comparative sampling by government agencies.

Private companies undertook about 2% of monitoring programs in 1998–99. Mining companies and a variety of industries undertook a large proportion of this monitoring as a requirement of their various licences.

A3.2.6. The Community

In its widest sense, the community represents both the ultimate source of water quality problems and the ultimate recipient of benefits flowing from their amelioration. To contribute to the long-term solution to water quality problems, the community must accept that it needs to modify existing behaviours. To facilitate this, the NWQMS and other initiatives have sought to engender an ethos of ‘ownership’ or participation among the community.

A role exists for local government, total catchment management committees, Landcare groups and other community-based organisations to encourage and foster wider community awareness and involvement in water quality monitoring and reporting. This is being done through programs such as Waterwatch. This involvement primarily educates the wider community but also

promotes its ownership of the monitoring program and the practical measures necessary to make key improvements.

It is preferable here to establish links with and to build successfully on existing public and community organisations in this field, rather than to try to develop new networks focused solely on water quality issues.

A3.3. Australian State of the Environment Reporting System

A3.3.1. Background

The Commonwealth State of the Environment Reporting system is an ongoing process for enhancing the quality, accessibility and relevance of data relating to ecologically sustainable development. The first major product of the system was *Australia: State of the Environment 1996* — an independent, nation-wide assessment of the status of Australia's environment. The recent *Environment Protection and Biodiversity Conservation Act 1999* now makes preparation of State of the Environment reports, and their tabling in Parliament, a legislative requirement. The Minister must cause a report on the environment in the Australian jurisdiction to be prepared in accordance with the regulations (if any) every five years (516B, EPBC Act 1999).

Arrangements are well under way for production of the 2001 Australian State of the Environment (SoE) Report, to be prepared by 31 December 2001 in accordance with the EPBC Act 1999. The 2001 SoE Report will concentrate on changes since the last report, cover new and emerging issues and pioneer the use of environmental indicators on a continental scale.

A3.3.2. Environmental Indicators for National State of the Environment Reporting

In the lead-up to the 2001 SoE Report the Australian State of the Environment Section (Environment Australia) has commissioned world-leading research culminating in the development of the series '*Environmental Indicators for National State of the Environment Reporting*'. The reports recommend indicators for each of the themes on which Commonwealth SoE Reporting is based: i.e. Human Settlements; Biodiversity; The Atmosphere; The Land; Inland Waters; Estuaries and the Sea; and Natural and Cultural Heritage. The advice embodied in these reports is also being used as an input to other initiatives, such as the National Land and Water Resources Audit (NLWRA) and the Australian Local Government Association's Regional Environmental Strategies.

The theoretical framework and sets of environmental indicators documented in these reports underpin the implementation of environmental indicators for the 2001 Report and future SoE Reports for Australia. Over the current reporting cycle, the Australian State of the Environment Section is working to secure access to relevant data for the 2001 SoE Report. For example, many data for environmental indicators for the State of the Environment reporting process are available from data sets managed by Commonwealth, state and territory agencies and data sets managed by Environment Australia.

Additional data for indicators will become available through existing sources and processes, and Environment Australia is continually looking to develop links with other agencies collecting and collating related data. The aim of these links is to maximise the effectiveness and efficiency of data gathering, particularly by avoiding duplication in establishing data sources. For instance, significant progress has been made between the Australian State of the Environment Section, the NLWRA and the states and territories with a joint water quality project that aims to: (i) implement environmental indicators related to exceedances of the water quality guidelines; and (ii) work towards the comparable analysis of water quality status and trends across all states and territories.

A3.3.3. Monitoring Needs and Strategies

The environmental indicators directly recommended for the Inland Waters theme are documented in Fairweather and Napier (1998) and indicators for the Estuaries and the Sea theme (now referred to as Coasts and Oceans) are documented in Ward et al. (1998). The reports identify suitable monitoring strategies for each indicator; including measurement techniques, appropriate temporal and spatial scales for measurement and reporting, data storage and presentation techniques, and the appropriate geographical extent of monitoring.

A3.3.4. Development of a Core Set of Environmental Indicators

The SoE environmental indicator series is being further complemented by a nationally agreed set of indicators, known as the 'Core Set', currently under development through the ANZECC State of the Environment Reporting Taskforce (ANZECC 1998). The Core Set represents those indicators that have been identified as useful for monitoring environmental trends at all spatial scales and therefore require consistent monitoring across jurisdictions. The endorsement of the core indicators would form an agreed framework for State of the Environment reporting across all state and territory jurisdictions.

Through this process, the core indicators will assist each jurisdiction to further develop its environmental monitoring, and will help to build a national picture of trends and the condition of our environment.

Further information can be obtained from the Australian State of the Environment Reporting Section (phone (02) 6274 1855).

A3.4. The Waterwatch Program

A3.4.1. General Information

Waterwatch Australia is a national network of more than 50 000 people who share a vision of healthy waterways. Waterwatch promotes water quality monitoring as a tool to involve the Australian community in land and water management at the local and catchment scales. Through monitoring their local waterways, communities are geared into action to address water quality issues and work together to protect and rehabilitate waterways.

Volunteer monitors identify problems and build up a long-term picture of the health of a waterway so that appropriate actions can be taken. Community-collected data can be used for a number of different purposes including teaching students about the importance of healthy waterways, as an early warning heralding particular water quality issues in a local waterway or wetland for local catchment planning, or as a contribution to State of the Environment reporting.

Waterwatch Australia is an umbrella program providing a national focus for state programs which include Waterwatch Victoria, Waterwatch South Australia, Waterwatch WA, Waterwatch NT, Waterwatch Queensland, Waterwatch NSW, Waterwatch ACT and Waterwatch Tasmania.

At present, more than 120 regionally-based coordinators are supported to varying degrees by Waterwatch Australia. These community employees are training others to get involved in Waterwatch and to 'read' the results of their monitoring so they can design projects to tackle the problems they detect. The Waterwatch network includes a State Facilitator in each state and territory and a National Facilitator in the National Office which is based in Environment Australia, the Commonwealth department for the environment, in Canberra. The Waterwatch Australia Steering Committee has also been established to provide direction and support for the program and to make program policy decisions on behalf of the network.

Since 1993, the Waterwatch Australia program has grown from 200 groups monitoring in 16 catchments, to nearly 1800 groups monitoring in more than 150 catchments. It is estimated that at present there are over 4000 Waterwatch sites being monitored by over 50 000 people across Australia.

Definition of the objective of monitoring and identification of the purpose of the data are both part of the Waterwatch monitoring plan. Potential data users include teachers, community groups, local governments, catchment management committees, industries and state agencies. Each of these users has different requirements of the data. The purpose for which the data will be used determines how reliable the data need to be and the required precision, accuracy and sensitivity of the equipment, and sampling and analysis procedures.

A number of resources and guidelines have been developed by Waterwatch Australia to assist community groups and individuals to collect credible, consistent and accurate data. Regional coordinators have been appointed to assist community groups to collect water quality data, to provide a focal point for the collation and interpretation of data and to facilitate the feeding of the collected information into local and catchment management planning processes. Regional coordinators provide training and support to the rest of the community. All the groups and individuals within a catchment are linked through the regional or catchment coordinator, ensuring that all data collected are interpreted in the context of the whole catchment. Coordinators provide regular feedback to volunteer monitors and other sectors of the community.

A3.4.2. Waterwatch Resources

The following resources assist the regional coordinators to provide support to the network.

(i) The Waterwatch Australia Manual (in preparation)

Title: *The National Technical Manual*

Availability: contact Waterwatch Australia

Description: The Waterwatch Australia Manual is a guide to monitoring for community groups and Waterwatch regional coordinators. The manual describes how to design a good quality monitoring program; how to monitor a large range of physical, chemical and biological parameters; quality control; how to interpret the results of Waterwatch monitoring; data management; and moving from monitoring to action.

(ii) The Waterwatch Australia Data Management system. This system has resulted from the need to integrate the range of data management and communication tools available to Waterwatch coordinators around Australia. A number of software tools have been developed to assist community groups to manage their data. These tools include:

- a Data Entry Program ensuring that all Waterwatch-collected data are recorded in a consistent manner across Australia. The Data Entry Program also provides a degree of validation and is available on CD-ROM or can be downloaded from the Waterwatch homepage, www.waterwatch.org.au;
- the Waterwatch Australia Database program, that enables regional networks to store, retrieve and report on their data; and
- GIS applications (state programs are currently developing tools to enable the Waterwatch network to link their water quality data to a GIS (geographic information system) to enable them to map, analyse and link information).

These tools may be used jointly or independently according to the needs of the individual coordinators.

The Waterwatch Offline data entry program is now available on www.waterwatch.org.au, the Waterwatch national web site.

Box A3.1. Draft Waterwatch sampling form, showing types of information that can be collected

Site Details Section

Site details are an essential part of the Waterwatch program. Accurate details of site locality — for example, grid references, catchment name and land use — are important for making comparisons between sites. This information can also be used to generate maps of the catchments to assist in further interpretation of results.

Group [.....], Contact [.....], Postcode [.....]

Sample date and time [.....]

State [..], Division [..], Basin [..], Catchment [..], Sub-catchment [..], Site name [..], Zone [..], Map no. [..] Altitude [..], Easting [..], Northing [..]

Land uses [..], Waterbody [..], Sample depth [..], Sample width [..], Rainfall [..]

Habitat Survey Section

A habitat survey involves visual assessment of the habitat value of the area immediately adjacent to your monitoring site. The condition of the vegetation in and around the waterbody provides a good indication of the likely condition of the aquatic environment. If the stream-side vegetation is intact, it provides a good buffer against erosion and the transport of sediment into streams and waterbodies. When stream-side vegetation is degraded, there is less protection against land use impacts and there may be subsequent deterioration of water quality.

Bank vegetation [..], Verge vegetation [..], Stream cover [..], Bank stability [..], Riffles, pools & bends [..]

Biological Monitoring Section

Biological sampling of water for quality focuses on macroinvertebrates and algae. Macroinvertebrates are central to the food chains in aquatic systems and are a good indication of stream health. They are sensitive to quite mild pollution or changes in water quality.

Macroinvertebrate Rating [....] *Diversity* [....]:

Beetles[..], Beetle larvae [...], Riffle beetles [..], Diving beetles [..], Water scavenger beetles [..], Whirligig beetles [..], Water pennies [..], Damselflies [..], Dragonflies [...], True bugs [...], Backswimmers [..], Water boatmen [..], Water scorpions [..], Water measurers [...], Water striders [...], Small water striders [..], Black fly larvae [...], Mosquitos [..], Chironomids [...], Biting-midges [..], Crane-flies [...], Stonefly larvae [...], Mayfly larvae [...], Caddis-fly larvae [...], Dobson-flies/Alderflies [...], Springtails [...], Water fleas [...], Clam shrimps [...], Yabbies/Marron/Gilgie [...], Freshwater shrimp [...], Freshwater prawn [...], Freshwater crab [..], Side swimmers [..], Freshwater slater [..], Water mites [..], Bristleworm [..], Leeches [..], Segmented worms [..], Snails [..], Limpets [..], Bivalves [..], Flatworms [..], Roundworms [..], Hydra [..].

Chemical and Physical Testing Section

Chemical and physical testing of water can provide an accurate analysis of the water in one particular place at one specific time. It gives a picture of the water composition and can confirm the presence of a particular type of pollution. Over an extended period, it can establish the normal levels for the waterbody and the typical range of values to be expected for the parameters measured.

Dissolved oxygen [..], % Dissolved oxygen [..], Water temperature [..], Velocity [..], Flow [..], pH [..], Electrical conductivity [..], Turbidity [..], Orthophosphate as P [..], Total P as phosphorus [..], Nitrate as N [..], Faecal coliforms [..], *E. coli* [..].

Other Information:

A3.4.3. Use of the Waterwatch Australia Data Management System

The three major components of the Waterwatch Australia Data Management System integrate the range of communication tools and data management techniques used by Waterwatch groups and coordinators around Australia. The package improves efficiency in the analysis and reporting of waterway monitoring information, it centralises the data at the regional scale, and it stores them in a uniform format.

The data are entered into four sections: site details data, habitat data, biological data and chemical and physical data (see Box A3.1). As well as results from monitoring, the data entered into the database include site location details (eastings/northings, latitude/longitude), group information, monitoring dates, frequency of monitoring, and sub-catchment/waterway information.

At present, statistical information is not generated at a national scale, although it may be generated at that scale in the future. The package can be used to compare, map, and graph data at a state and regional scale depending on how many regional coordinators are using the package to store data.

Data have been used by some groups to generate catchment/waterway reports that in some cases are printed in the local newspapers for general community environmental awareness.

A3.4.4. Preparation of Monitoring Plans

Waterwatch monitoring groups are encouraged to consider a range of issues when preparing a monitoring plan. The following questions are used to help a group to design an effective monitoring plan:

- why is the group monitoring a particular waterway or site?
- who will use the monitoring data?
- how will the data be used?
- what parameters will each group monitor?
- what level of reliability of data does each group want?
- what methods for data collection will the group use?
- where will each group monitor?
- when and how often is the group able to monitor?
- who will be involved and in what way?
- how will the data be managed and reported?
- what quality control measures will be used to ensure accuracy, reliability and credibility?

The Waterwatch Australia Manual discusses the processes for developing monitoring plans and the issues that need to be considered by monitoring groups.

A3.4.5. Unique Site Codes

To ensure that Waterwatch data are able to be stored and retrieved at a state and national level, a unique site code system has been developed. The unique site code system is based on the Australian Water Resources Commission drainage divisions and basins and comprises an AWRC basin, a three-letter catchment code and a three-digit number. Every Waterwatch monitoring site in Australia has a unique site code.

A3.4.6. Quality Control

Waterwatch data are being used for local community action plans, local government planning processes, catchment management planning, environmental impact statements and as baseline information in catchments where there have never been monitoring programs. The community has

developed increased skills and knowledge in monitoring procedures, and both the data collectors and the data users are placing greater demands on the data to be accurate and reliable.

In response to this demand Waterwatch has developed a Draft Data Confidence Plan that provides protocols and procedures to enable the community to collect data that are of a level of quality for the purpose for which they are to be used.

A3.4.7. Waterwatch Contact Details

Details of Waterwatch Contacts are given in Table A3.2. Web site addresses are also provided for the state/territory programs. These web sites provide technical information and forums where individuals and groups can share information and seek advice from others in the network. Sharing of information throughout the network should enhance the quality of the information collected about individual waterways.

Table A3.2. Waterwatch contacts

National Facilitator
Kate Gowland
Environment Australia
GPO Box 787, Canberra ACT 2601
ph: 02 6274 2797; fax: 02 6274 2735
email: kate.gowland@ea.gov.au
www.waterwatch.org.au

New South Wales
Geoffrey Smith
Dept of Land & Water Conservation
GPO Box 39, Sydney NSW 2001
ph: 02 9228 6571; fax: 02 9228 6464
email: gsmith@dlwc.nsw.gov.au
www.streamwatch.org.au

Australian Capital Territory
Jane Horniblow
Environment ACT
PO Box 144, Lyneham ACT 2602
ph: 02 6207 2246; fax: 02 6207 6084
email: jane.horniblow@act.gov.au
www.act.waterwatch.org.au

Tasmania
Michael Cassidy
Dept of Primary Industry, Water & Environment
PO Box 46, Kings Meadow TAS 7249
ph: 03 6336 5254; fax: 03 6336 5311
email: Michael.Cassidy@dpiwe.tas.gov.au

Victoria
Vera Lubczenko
Dept of Natural Resources & Environment
PO Box 41, East Melbourne VIC 3002
ph: 03 9412 4663; fax: 03 9412 4039
email: Vera.Lubczenko@nre.vic.gov.au
www.vic.waterwatch.org.au

South Australia
Ross Wissing
Environment Protection Agency
Dept for Environment and Heritage
GPO Box 2607, Adelaide SA 5001
ph: 08 8204 9117; fax: 08 8204 9761
email: rwissing@deh.sa.gov.au
www.sa.waterwatch.org.au

Northern Territory
Leslie Alford
Natural Resources Division
Dept of Lands, Planning and Environment
PO Box 30, Palmerston NT 0831
ph: 08 8999 4456; fax: 08 8999 4445
email: leslie.alford@nt.gov.au
www.lpe.nt.gov.au/waterwatch

Queensland
Christina Dwyer
Dept of Natural Resources, Resource Science Centre
Gate 4 Block 80 Meiers Rd, Indooroopilly QLD 4068
ph: 07 3896 9737; fax: 07 3896 9625
email: Christina.Dwyer@dnr.qld.gov.au
www.qld.waterwatch.org.au

Western Australia
Julie-Anne Smith
PO Box 6740, Hay St
East Perth WA 6872
ph: 08 9278 0359; fax: 08 9278 0301
email: julie-anne.smith@wrc.wa.gov.au
www.wrc.wa.gov.au/ribbons

A3.5. Future Directions

A3.5.1. Coordination

Reducing Overlap

There is potential for overlap and duplication between the agencies which undertake monitoring. Some states have lead agencies to coordinate monitoring, and statewide databases are being developed. For an example, now available on the Internet, see section 7.3 (page 7-3).

Moves towards regular wide reporting of monitoring results should be encouraged, including expansion of existing internal reporting towards future 'State of Rivers' reports, leading in time to State of the Environment reports. It is expected that this kind of reporting, in itself, will generate a requirement for specific kinds of data.

Development of Protocols

A number of bodies have publicly available protocols, but it is not clear how universally they can be applied across Australia. The importance of consistency of reporting and comparability of data dictates that consistent or standardised protocols should be developed. These Monitoring Guidelines are a useful start in that direction. DEST (1995) found that, although there were many monitoring programs, some of the data collected so far had not provided adequate information for ongoing management of water resources (i.e. the 'data rich–information poor' syndrome), and there were still significant gaps in water quality information which needed to be addressed.

This is a major task though it is possible to confine it by only having protocols for certain key indicators. A mechanism is needed for agreeing on prioritised standard protocols and the ongoing adoption of such protocols.

A3.5.2. Future Directions in Monitoring Technology

Future programs will make greater use of continuous electronic remote monitoring equipment that can indicate physical and chemical changes that affect aquatic organisms and the surrounding biota. These techniques will need to be tuned to specific water quality monitoring needs.

Developments in monitoring technology are progressing towards continuous, automated sampling systems that are sensitive to a broad range of chemicals and can be monitored remotely using satellite technology. Sensor technology and sampler design and the establishment of uniform monitoring and equipment guidelines will result in analytical data of high validity and quality.

If data acquired by on-site water samplers can be available by remote access, the labour costs involved will be reduced, and accurate real-time data will be obtainable from isolated areas.

Some significant recent developments in terms of water quality monitoring technology include:

- the CSIRO developed 'Sewer Sentinel' which is capable of continuously monitoring the temperature, pH, conductivity, dissolved oxygen, and turbidity of raw sewage. Sewer Sentinels have been extensively trialled in a range of domestic and industrial sewers and have shown great potential.
- a device for the real-time monitoring of dissolved phosphate in final tertiary-treated plant effluent, developed jointly by Monash University Water Studies Centre, University of NSW Centre for Membrane Science and Technology, the Science and Environment Division of Australian Water Technology, and BHP Research;
- an automated analytical device developed by CSIRO and commercialised by Greenspan that uses physical, electrochemical and spectrophotometric sensors to measure a wide range of variables simultaneously. The unit is solar powered, features a highly accurate automatic calibration system and built in logger, and can be accessed remotely via satellite.

- research by La Trobe University into electrochemical sensor technology. Projects include the development of a voltametric analytical instrument for the simultaneous measurement of K, Ca, NO_3 and Cl ions in waste water.
- an acoustic Doppler current meter developed by the University of Western Australia which measures water flow with great accuracy.

Standard approaches to sampling, analysis and statistics will enable increased use of data from a range of monitoring programs by different organisations.

Areas that have been targeted for further development include:

- sensors that can provide greater detail about particular forms of contaminants and that can detect lower concentrations of toxic organic compounds in natural waters;
- new methods for the handling and processing of 'dirty' and heterogeneous samples;
- real-time methods for the assessment of microbial quality of waters; and
- a design for a suitable platform to house and protect the electronic sensors used for continuous measurement. The huge range between wet and dry in some rivers poses problems in terms of securing the equipment and sourcing units for long-term operation.

The main challenge in manufacturing real-time monitoring equipment for remote areas is to produce instruments that are as reliable, robust and low in maintenance as possible. These instruments either should not require frequent calibration or should perform self-calibration as often as required.

A3.5.3. Biomonitoring

Future surface water quality monitoring will use the latest biological monitoring techniques and directly measure changes in population dynamics that can show the effects of combined environmental stresses and inefficient management practices. 'Low-technology' monitoring techniques such as biomonitoring are attracting increasing interest because of the need to combine physical, chemical and biological information to make meaningful environmental interpretations.

Biomonitoring measures the responses of aquatic organisms to contaminants and environmental conditions. The data obtained from biomonitoring programs are useful in conjunction with physical and chemical data. Macroinvertebrates are most commonly used because they are easy to identify. The group comprises a wide range of organisms and therefore a variety of responses can be observed; their sedentary habits are conducive to spatial sampling and their relatively long life spans permit temporal changes to be observed.

Biomonitoring techniques have been used for many years in the UK, USA and Canada, but Australia has some unique characteristics in contrast to other developed countries, such as the highly variable but generally low flow, and the time of peak detritus input. Therefore, these techniques need further development for Australian conditions. Research in Australia into rapid bioassessment techniques, based largely on macroinvertebrates, is being undertaken by various institutions and individuals including Victorian agencies, the Cooperative Research Centre for Freshwater Ecology, and groups associated with the National River Health Program (NRHP). The NRHP is focusing on macroinvertebrates but also considering algae, diatoms, macrophytes and fish.

Rapid bioassessment techniques reduce the time and costs involved in the assessment of environmental conditions, and facilitate the presentation of data that can be easily interpreted and understood.

Appendix 4
Water Quality Monitoring Case Studies

A4.1. Case Study 1: Eutrophication of the Upper Murrumbidgee River and Burrinjuck Reservoir

A4.1.1. Objectives and Information Requirements

Following earlier water quality and ecological investigations of the Upper Murrumbidgee River and Burrinjuck Reservoir an extensive ongoing monitoring program was initiated of the Murrumbidgee River, downstream of Canberra. The study objective was to monitor the water quality and aquatic ecology of the Murrumbidgee River and its associated catchment in relation to efforts to ameliorate eutrophication in the Upper Murrumbidgee River and Burrinjuck Dam.

The following information requirements were identified at the beginning of the study:

- What is the nature and extent of nuisance algal growths in the river and Burrinjuck Reservoir?
- What are the current environmental and use values of downstream waters, and what are the water quality guidelines consistent with meeting these values?
- What are the critical contaminants causing the nuisance algal growth, and what is the sustainable level of loading consistent with restoring acceptable levels of water quality?
- What are the major sources of nutrients causing the nuisance algal growth?
- What is the role of climate (catchment runoff) in modifying the levels of nutrients discharged and the responses of the receiving water?
- How should the new Canberra sewage treatment facility be operated to minimise the problem?
- What are the management practices in respect to other land uses across the catchment?

A4.1.2. Study Design

Conceptual Model

These were the initial assumptions or conceptual models interrelating pressure and state conditions.

- Algal biomass levels in Burrinjuck Reservoir depend on the annual external loading of phosphorus, and water loading or hydraulic residence time.
- A substantial mass of nutrients is contributed by the catchment during non-point source discharge events: both non-point source event discharges and point source discharges during periods of low flow are important.
- Contaminant levels gradually decay downstream of discharge points as a result of biological uptake and oxidation of waters.

More recently, there has been an improved understanding of nutrient pathways and transformation processes. When water enters the system during rain events, high in suspended solids, the nutrients are rapidly adsorbed onto suspended particles and removed from the water column by sedimentation. During low flow conditions, when the water has relatively few suspended solids, ex-sewage nutrients may be taken up by algae, either attached or planktonic. The algae ultimately die and settle to the sediments, imposing a significant loading of organic carbon on the sediments. Subsequently, sedimented phosphorus is made bioavailable during the

chemical reduction of sediments and remobilisation of phosphorus as orthophosphate. Sediment reduction is driven by the loading of organic carbon (BOD) and the mixing conditions (oxygen transfer rates offsetting the BOD) for the particular waterbody; that is, discharged phosphorus may be only indirectly involved in driving the in-stream remobilisation of sedimented phosphorus and subsequent algal growth.

Study Boundaries

The monitoring area covers the catchment and streams upstream of Canberra, and the catchment area and streams downstream to and including Burrinjuck Reservoir.

Measurement Parameters

To respond to study objectives, it was necessary to monitor:

- the mass of contaminants discharged from the catchment during non-point source discharges, expressed as a function of depth of runoff, and land use and management practice;
- the mass of contaminants discharged from point sources, expressed as a function of discharge rate and treatment facility;
- the transport losses by sedimentation, or the gains by re-suspension or microbial remobilisation, as a function of distance downstream from the discharge point, travel time, and flow rate or reach loading;
- the algal response to the composition, concentration or load of nutrients, and mixing and light conditions;
- the modification to algal composition and biomass as a result of zooplankton grazing.

Table A4.1 lists the parameters that were recorded during the study. It was important to identify the specific variables that were used to address the conceptual model of the study. Phosphorus concentrations were a major theme within this study because this nutrient's concentrations were related to algal counts and were also used to identify the point source and non-point source contaminant sites within the catchment. Another variable that interacted with the relationship between phosphorus concentrations and algal counts was turbidity. Apart from the external sources of phosphorus, it was possible that the lake sediment would need to be analysed. All these types of relationship needed to be determined as part of the study design.

Table A4.1. List of parameters sampled in streams and lakes

Type	Parameter
Physical	Turbidity, temperature, colour, electrical conductivity, total dissolved salts, total suspended solids*
Chemical	pH, dissolved oxygen, alkalinity, ammonia, nitrate+nitrite, organic nitrogen*, total phosphorus, orthophosphate, dissolved organic carbon, chloride[88], silica[88], sodium[88], calcium[88], magnesium[88], iron[88], manganese[88], potassium[88], zinc[88], copper[88], nickel[88], sulfate[88], chlorophyll-*a*
Biological	Algal cell numbers and composition, zooplankton numbers and composition, *E. coli*

* Included since 1983; [88] Occasional analysis only

Sampling Sites and Frequency

The selection of the monitoring sites, sampling frequency and selection of analytes was guided by the following principles:

- the need to monitor loads upstream and downstream of Canberra, including in-stream decay in loads downstream of discharge points;

- the need for event-based monitoring, in addition to routine monitoring of point source discharges;
- the need for special studies to better describe in-stream physical, chemical and biological processes:
  – analysis of in-stream mixing and diffusion processes, using dyes;
  – role of stratification within the Burrinjuck Reservoir (depth profiles);
  – biological composition of streams (macroinvertebrate sampling, fish surveys);
  – nutrient spiking of macro-enclosures to assess algal responses to different nutrient regimes;
  – analysis of nutrient release from Burrinjuck sediment cores under anaerobic conditions;
- the need for an initially intensive sampling frequency (5-day), to see how frequent the sampling had to be to adequately assess variability in the system;
- the need to target both phosphorus and nitrogen as possible critical contaminants;
- the need to select sampling points that provided representative (well mixed) samples and were linked with stream gauging information;
- the need to adopt auto-samplers as the only viable basis for obtaining event samples in urban catchments;
- the need to adopt a range of biological surveys to assess the health of the system.

In view of the contaminant pathways and transformation pathways unique to each sub-system, and their individual temporal and spatial variability properties, it was necessary to address sampling design at least at the sub-system level.

In the case of the Murrumbidgee and Burrinjuck system, there were four distinct sub-systems:
- rural catchment and streams upstream of major sewage effluent discharges;
- urban catchments and streams;
- major streams downstream of major sewage effluent discharges (Canberra, Queanbeyan);
- large downstream lakes (Burrinjuck Reservoir, Lake Burley Griffin).

The location of the sampling was also a factor that had to be considered. It was important to gain representative samples of the sub-systems being studied. The frequency of the sampling was highly related to the sub-systems' hydrological characteristics. The events that occur within an urban environment take place very rapidly with high energy outputs, compared to upper reaches in which events may still be detected weeks later downstream.

Assessment of required sampling frequency:
- in-stream water quality (non-event conditions)
  0–3% error in site median values for 14-day-based sampling versus 5-day;
- in-stream water quality (event conditions)
  8-hour-based sampling for primary streams, 20-minutes-based sampling for urban streams and drains;
- in-lake water quality
  3–8% error in site median values for 14-day-based sampling versus 7-day.

A4.1.3. Sampling and Analysis

Field Measurements

The turbidity, temperature, colour, conductivity, pH and dissolved oxygen were all measured in situ. All the electronic equipment was calibrated before sampling. The in-field staff that performed the sampling were trained in the appropriate sampling procedures. Alkalinity was determined via a simple in-field titration using dilute acid and methyl indicator.

Laboratory Analysis

All the analyses performed within the laboratory followed standard methods that are specified by *Standard Methods for the Examination of Water and Wastewater* (APHA 1998) or USEPA sampling and analysis methods (Keith 1991). As part of the analysis the samples analysed were documented in detail. The laboratory facilities and analytical equipment were maintained to a very high standard to ensure minimal contamination. All staff that performed the analysis were competent and followed up the analysis by documenting the analytical methods used, and using appropriate QA/QC.

Data Management

There was early recognition of the importance of establishing an archive for water quality data, with checking and validation protocols for the entry and management of data, and documentation of sampling and analytical techniques and methods.

In view of the volume of data and the range of groups participating in sampling and analysis, only by this means could the monitoring program ensure:

- systematic logging, validation and entry of data;
- secure storage;
- consistent nomenclature, procedures and analytical methods;
- ease of data access.

Entries into the archive comprised information on the sample data, field observations or comments, sampling methods, site descriptions and river flows at the time of sampling. User-specified reports were based on site description (number), date and parameters required, and the laboratory undertaking the analysis.

For quality assurance/quality control,

- validation checks were standard for various fields in the database to minimise data entry errors;
- data analysis and interpretation were peer reviewed.

A4.1.4. Data Analysis and Interpretation

The initial data analysis comprised statistical analysis of medians, ranges, trends (time and distance downstream), and correlation analysis of the parameters.

The explanatory power (validation) of a range of conceptual models was tested; the models included:

- in-stream steady-state first-order time-based decay of contaminants;
- in-stream time-series-based analysis (Centre for Resources and Environmental Studies CAPTAIN package);
- in-stream load, sedimentation and remobilisation or re-suspension-based models (AQUALM);
- in-lake load- and retention-time-based algal response relationships (Vollenweider).

A range of restoration measures was adopted as part of an integrated catchment management strategy. They included the tertiary treatment of Canberra sewage (98% removal of phosphorus, 95% removal of nitrogen, and 98% removal of phosphorus alone), and urban stormwater pollution control ponds, gross pollutant traps and ‘at-source’ controls (80% removal of suspended solids and 70% removal of total phosphorus).

Since the restoration measures were set up, the monitoring analysis has included:

- time-based trend analysis;
- comparison of average summer algal levels, normalised for varying streamflow or water loading conditions (modified loading-based model);
- changes in biological indices of river health (AUSRIVAS).

Currently, the Burrinjuck physical mixing conditions, external and internal loading, and algal biomass and composition are being reviewed for the 22 years of data (National Eutrophication Management Program: Burrinjuck Algal Succession Project).

A4.1.5. Reporting

Reporting has comprised Project Reports, annual reviews of water quality, and web-based information. Access to the water quality database is available through the World Wide Web.

A range of techniques have been used to present the data analysis, including:

- time-based trend lines of water quality;
- box and whisker plots (medians, 90th and 10th percentile values) for a range of sites and determinands;
- charts based on nutrient and water load – algal response, for the major lakes and reservoirs;
- River Health Indices for a range of sites.

A4.2. Case Study 2: A Groundwater Quality Assessment of the Alluvial Aquifers in the Logan–Albert Catchment, SE Queensland

A4.2.1. Objectives and Information Requirements

Natural resources in the Logan–Albert Catchment are being subjected to increasing pressure as a direct result of population growth in South-East Queensland. This has prompted an investigation of the aquifer that supplies this region with its groundwater. The study was initiated in June–July 1994.

The major objective was to establish a groundwater quality condition benchmark for use in subsequent monitoring, to identify and understand the processes degrading groundwater quality in the aquifer, and to integrate the information obtained and provide advice to the responsible natural resource managers.

These information requirements were identified during the design phase:

- What is the geology of the Logan–Albert catchment?
- What are the hydrological characteristic of the three tributaries in the catchment, the Logan River, Albert River and the Teviot Brook?
- What are the existing physiography, climate and soils of the areas under investigation?
- What land uses exist within the catchment and is there a correlation between the use and physiographic features?
- What is the current, and past, water consumption within the catchment?
- What are the aquifer yields within the catchment, and does this relate to the aquifer type, i.e. unconsolidated sediment, sedimentary rock, fractured rock?
- What is the current water quality, in relation both to human consumption and irrigation quality criteria, of the identified aquifer types?
- Do water and sewage treatment plants exist within the catchment, and do these plants affect the aquifer water quality?

A4.2.2. Study Design

Conceptual Model

The initial assumptions and conceptual models interrelating pressure and state conditions were:

- aquifer water quality is related to land use practices within the catchment;
- the geology of the catchment directly relates to the physical characteristics of the alluvial aquifer;
- increasing land use pressures are leading to poorer groundwater quality.

Before this study, there had been very few projects that focus on groundwater within the Logan–Albert catchment. The main reasons stated by Please et al. (1997) for this were:

1) the primary source of public supply water is from surface water. The main use of groundwater is for irrigation and private supplies for farms.
2) this region is an undeclared zone, so the government has no statutory power to license private bores. Therefore, there is no requirement for private bore data to be passed on to government bodies for storage and use in investigations.

The major theme in early investigations had related to surface water quality; groundwater was neglected. It was not until 1979 that an attempt was made to investigate groundwater quality. This investigation, initiated by Quarantotto (1979) (cited in Please et al. 1997) concluded that the Logan–Albert region had poor groundwater quality which was potentially due to the poor yields and/or poor water quality in the main hydrological provinces. In subsequent years several projects were undertaken to determine groundwater quality (Please et al. 1997), focusing mainly on in-fill information, inorganic constituents including trace metals and nutrients, and analysis of a few trace organic compounds and pesticides. Throughout all these investigations the Quaternary alluvial aquifer was the major theme. This system supplies and regulates the groundwater within the catchment. The geological characteristics of a given area will directly affect the groundwater quality.

Study Boundaries

The study was confined to the Logan–Albert catchment.

Measurement Parameters

To respond to the issues identified, the following monitoring information was required:

- water yields from the Quaternary alluvial aquifer, based on the location within the Logan–Albert catchment;
- assessment of physical parameters, inorganic chemistry (including metals and environmental isotopes), nutrients, contaminant and indigenous microbes, and pesticides within the groundwater of the Logan–Albert catchment;
- assessment of the potential sources of contaminants to the groundwater system. This was based on identifying specific areas and monitoring them in relation to their output of hydrological parameters.

The design of the sampling sites, sampling frequency and selection of analytes was guided by the following principles:

- the need to choose bore sites for groundwater quality sampling based on the existence of bores and associated data, availability, accessibility, bore diameter, flow rates and general quality, and spatial coverage;
- the need to pre-test cores to identify suitable study cores;
- the need to place bore lines strategically within the catchment to gain a representative indication of each individual river system's association with the aquifer;
- the need to establish a groundwater quality condition benchmark within the Logan–Albert catchment, based on ambient concentrations of identified hydrological parameters and the physical condition of each site;

- the need to identify and understand processes affecting groundwater quality in the aquifer;
- the need to integrate information obtained and provide data that would be useful to resource managers;
- the need for special studies to better describe chemical and biological processes in the aquifer:
  – the fate of contaminants after they enter the aquifer;
  – the effects of excessive nutrient input on aquifer water quality, biological and chemical perspectives;
  – the physical characteristics of the aquifer that would promote the entry of contaminants;
  – contaminant spiking to assess the transportation pathways of contaminants within the aquifer;
  – assessment of the temporal and spatial characteristics of contaminants within the aquifer.

The following measurement parameters were selected:

(i) major and minor inorganic constituents:

- sodium, potassium, calcium, magnesium, sulfate, chloride, bicarbonate;
- trace elements and metals;
- alkali metals — lithium;
- alkaline earth metals — barium, strontium;
- metallic elements — aluminium, cadmium, chromium, cobalt, copper, gold, lead, mercury, molybdenum, nickel, silver, tin, vanadium, zinc;
- non-metals — antimony, arsenic, boron, selenium;
- radioactive elements — uranium;

(ii) nutrients and carbon:

- an analysis of nitrate species for each location was performed, and ammonium-N was determined in conjunction with this analysis;
- phosphorus/phosphate concentrations;
- dissolved organic carbon;
- herbicides including triazines;
- insecticides including organochlorines;
- fungicides including conazoles;

(iii) microorganisms:

- faecal indicator bacteria;

(iv) environmental isotopes:

- tritium;
- deuterium oxygen;
- chlorine-36.

A4.2.3. Sampling and Analysis

Sampling Sites

In all, 36 (QDNR) observation bores were sampled along 17 borelines, with 12 bores being analysed from each catchment during the study period (Table A4.2). The data required included physical parameters, inorganic chemistry (including metals and environmental isotopes), nutrients, contaminant and indigenous microbes, and pesticides. Potential sources of contamination to the alluvial aquifer from human activities include agricultural and rural industrial land uses, particularly from beef and dairy production. Within the Logan–Albert catchment there are many pre-existing bores that are potential sampling locations. Based on this selection of bores, sampling locations were chosen that would best represent the three individual catchments. Before any sampling, pre-testing was undertaken to establish which bores were suitable for analysis.

Table A4.2. Groundwater bores drilled, tested and sampled

	Logan Catchment	Albert Catchment	Teviot Brook Catchment	Total
Drilled	63	44	25	132
Pre-tested (May 1994)	15	13	none — all considered to be low risk	28
Sampled (June–July)	12	12	12	36

Note: one surface water sample taken from Logan River next to Round Mountain bore line

Field Measurements

Five field measurements were taken at each bore:

1. total dissolved solids (TDS),
2. pH,
3. electrical conductivity (EC),
4. redox potential (Eh),
5. dissolved oxygen (DO).

The quality assurance and quality control guidelines relating to field data collection were as follows:

- adhere to standard sampling protocols;
- use correct sampling equipment, and transport and store samples in the appropriate manner;
- ensure that all equipment is calibrated before use.

Laboratory Analyses

The quality assurance and quality control guidelines relating to laboratory analysis were divided into three categories.

(i) Blanks were used to monitor contamination during any stage of the sampling and analytical process. Blanks were taken at the beginning of the trip, mid-way and at the end of the trip. A 'before' and 'after' blank was taken at each time, i.e. the distilled water before and after it had been through the decontaminated pump system. Separate blank samples were used for chemical, microbiological and isotopic analyses. 'Before' and 'after' blanks were taken at more frequent intervals for microbiological samples.

(ii) Duplicate samples were taken as a test of precision in sampling and analysis. They were taken every tenth sample and processed in exactly the same way as the 'normal' samples.

(iii) Standard additions were used to test the accuracy of the analytical instruments. In this project the samples were spiked in the field to determine degradation between collection and analysis. A spiked sample was prepared every time a duplicate sample was taken. Spiked samples were prepared for major and minor ions, metals and nutrient analyses.

Pesticide analysis involved specific QA/QC procedures.

(i) Blanks were collected on three occasions throughout the sampling period. These ensured that the rinsing procedure between sampling was efficient and that there was no contamination. 'Before' blanks were performed by extracting pesticide-free water. 'After' blanks were performed on pesticide-free water which had passed through the pump equipment after the routine cleaning of the equipment with detergent, pesticide-free water and analytical methanol.

(ii) Duplicate samples were taken on three occasions for the normal extractions. The duplicate results monitored precision of sampling, extraction and analytical methods.

(iii) Triplicate samples were spiked with a range of compounds, representing those pesticides of interest. These recoveries gave an indication of the extraction efficiency of a range of compounds, matrix effects and degradation of analytes with storage. Acidified cartridges were also spiked on three occasions with a mixture of acid herbicides.

(iv) Each sample was spiked with a known quantity of surrogate solution before extraction. This solution consisted of the following compounds: difluoro-DDE; dibromo-DDE; deuterium-labelled terphenyl (δ_{14}). The recovery of these compounds indicates the efficiency of the individual cartridge extraction.

One grab sample was taken at the Logan River. This site was chosen to detect possible surface–groundwater interactions with the closest bore sample.

Data Archive

There was early recognition of the importance of establishing a water quality data archive, including checking and validation protocols, for the entry and management of data, and documentation of sampling and analytical techniques and methods.

In view of the volume of the data and range of groups participating in sampling and analysis, only by this means could there be:

- systematic logging, validation and entry of data;
- secure storage;
- consistent nomenclature, procedures and analytical methods;
- ease of data access.

Entries into the archive comprised information on the sample data, field observations or comments, sampling method, site description and condition at the time of sampling. User-specified reports were based on site description (number), date and determinands required, and the laboratory undertaking the analysis.

For quality assurance/quality control:

- validation checks were standard for various fields in the database to minimise data entry errors;
- data analysis and interpretation were peer reviewed.

A4.2.4. Data Analysis and Interpretation

The interpretation of the data was based on four main objectives:

- to set a benchmark for groundwater quality for 1994. This involved clear display of the analytical results for quick assessment and characterisation of groundwater quality.
- to understand the hydrological processes of the region through hydrochemical evaluation;
- to define any zones that could be affected by human activities;
- to provide information to assist in the sustainable management of natural resources.

With these guiding objectives in mind the interpretation of the data could be undertaken.

Several specific types of data evaluation methods were employed to illustrate and better understand what the data were representing:

- bar graphs of hydrochemical parameters against core number,
- spatial distribution of key parameters graphed on a schematic diagram,
- *X–Y* graphs for selected parameters,
- piper triliner diagrams for major ions,
- schematic diagrams showing isotopic concentrations.

A4.2.5. Reporting

All field parameters were presented in bar graphs. Incorporated into these was a comparison of the in-field measurements and laboratory measurements. All the major and minor inorganic constituents' results were presented in bar charts, and also graphed against total ion concentration to identify any correlations. Sulfate concentrations were represented in *X–Y* graphs against dissolved oxygen, to identify any correlation between these two parameters. As well, iron and manganese concentrations were graphed against dissolved oxygen, Eh, pH and SO_4.

Trace element and metal parameters were all presented in bar graphs. Also, there was regression analysis of strontium and calcium concentrations and of aluminium and silicon. A graph of vanadium vs. Eh was also included in this section.

Nutrient concentrations were presented as bar graphs, with comparison between selected parameters. Nitrate was graphed against ammonium concentrations. Nitrate was also graphed against electrical conductivity, Eh, dissolved oxygen, depth of sample and depth to potentiometric surface (with and without removal of 'non-detects'). Orthophosphate was graphed against total phosphorus, and total phosphorus was also graphed against total nitrogen. Dissolved organic carbon was graphed against manganese, iron and sulfate.

Faecal indicator bacteria (FIB) concentrations (colony forming units, CFU) were presented in a table, with dissolved organic carbon. A schematic map of the FIB contamination in the individual catchments was also provided.

The environmental isotope concentrations of tritium, chlorine and deuterium were all presented in bar graphs. They were also displayed in schematic maps showing their activity in the catchments. Tritium was also graphed against dissolved oxygen, Eh, bicarbonate or total major anions and nitrate. Tritium and chlorine were graphed against each other for all catchments.

A4.3. Case Study 3: Water Pollution in the Derwent Estuary, Tasmania

A4.3.1. Objectives and Information Requirements

The Derwent Estuary in Hobart has sustained extensive organic, heavy metal, nutrient and pathogen pollution since the 1920–30s, from paper processing, zinc processing and sewage discharges. The contaminants have disturbed the ecology of the estuary, and its commercial fishery and recreation. There are also concerns about the effects of introduced marine organisms on local estuarine ecology.

The Department of Environment and Land Management (now E&P/DELM) established a variety of medium- to long-term monitoring programs, covering toxicants, nutrients, algae and bacteria, with the objective of ascertaining the current environmental health of the estuary. In addition, the Department has required annual surveys of heavy metals in aquatic organisms (by Pasminco Hobart Smelter), since 1972.

The objectives driving the monitoring and related studies have been:

- to determine the water quality guidelines appropriate for protecting and restoring the environmental values of the estuary;
- to determine the range of measures necessary to restore the water quality and ecology of the estuary, and the priorities for implementing the measures;
- to determine the appropriate emission control standards.

Currently, the monitoring studies have the objective of documenting existing conditions, tracking long-term trends, and determining if the estuary waters are suitable for primary contact recreation.

A 'State of the Derwent Estuary' review (Coughanowr 1997) was undertaken to identify significant pollution sources and to evaluate proposed remedial works. Information from various monitoring programs, environmental quality investigations and modelling studies from 1970 to 1997 was gathered and summarised to:

- identify and quantify (where possible) major sources of pollution and pollution loads to provide a '1996 snapshot';
- identify, compile and review existing data on water quality, sediments and biota prior to 1997;
- inform and educate resource managers and the public;
- identify gaps in the existing information database; and
- establish benchmarks for determining trends and improvements in the environmental quality of the estuary.

This is an example of a study that aims to develop a more focused water quality monitoring program by examining existing information and highlighting objectives, management issues, water quality concerns, monitoring and sampling design limitations and environmental processes.

A4.3.2. Study Design

Conceptual Model

These were the initial assumptions or conceptual models interrelating pressure and state conditions:

- This is a highly dynamic system, with tidal exchange, variable inflows from the Derwent River, pulses of industrial organic and heavy metal wastes, and sewage discharges high in nutrients, organic material and pathogens.
- There is substantial diurnal mixing and diffusion as a result of tidal exchange in the lower reach, with salinity stratification of waters in the middle and upper reaches, exacerbating pollution processes.
- Sediments hold significant stores of contaminants and are sources of contaminants as a result of remobilisation and re-suspension.
- Reductions in the contaminant loads will yield improved water quality and the recovery of ecology.

Study Boundaries

The monitoring extends from the Derwent River at New Norfolk to the Derwent Estuary discharge into Stormwater Bay.

Measurement Parameters

The following measurements were required, to respond to the information requirements:

- the mass of contaminants discharged from the catchment during non-point source discharges;
- the mass of contaminants discharged from point sources;
- the mass of contaminants introduced by tidal inflows, including variability in Storm Bay nutrient levels;
- the mass of contaminants discharged in tidal outflows;
- the contaminant transport and transformation pathways for a range of mixing conditions;
- the internal loading on the estuary (sediment reduction or re-mobilisation rates);
- the oxygen budgets and relationship to sediment BOD -de-oxygenation in stratified zones;
- the concentration and bioavailability of toxicants as a function of direct and indirect loading, and their effects on biota;

- the algal response to the composition, concentration or load of nutrients, and to mixing and light conditions;
- the seagrass response to increases in suspended solids and dissolved nutrient–epiphytic density.

Specific measurement parameters selected are shown in Table A4.3. The table also shows the parameters being measured in the most recent studies, initiated in 1999.

A4.3.3. Sampling and Analysis

Monitoring programs have been severely constrained by budgetary limitations. Consequently, it has not been possible to intensively monitor the following issues:

- non-point source load conditions during major river discharges;
- tidal inflows and outflows over tidal cycles;
- transport and transformation pathways along the estuary.

Monitoring has been largely limited to routine sampling of water quality through the estuary, as the basis for establishing the water quality prevailing as a result of current loading regimes and tidal exchange and river inflow conditions.

Knowledge of tidal exchange and mixing conditions is fundamental to understanding the dominant pollutant transport and transformation pathways. Only relatively recently has there been an attempt to develop a mixing and tidal exchange model for the estuary, on which a more integrated and systematic monitoring program might be based. A number of studies have been conducted to understand the circulation and mixing of the Derwent Estuary, including the recent study under the CSIRO Coastal Zone Program.

The design of the monitoring sites, sampling frequency and selection of analytes was guided by the following principles:

- the need to monitor discharges of industrial and municipal waste waters to the estuary;
- the need to monitor water quality of estuary waters at sufficient frequency and sites along the estuary to describe the prevailing water quality;
- the need to periodically survey sediments to assess net increase or loss in pollutants;
- the need to sample a range of biota to assess toxicant uptake rates and levels;
- the need to periodically survey biota (including seagrass, algae, benthic macroinvertebrates, fish) to assess changes in the ecological impacts.

A variety of water quality monitoring programs have been established. There have been semi-annual (2–5 times per year) water quality surveys of physical determinands, heavy metals and faecal bacteria (by E&P/DELM), during 1971–88, 1993–98 at 14 sites. There have been annual surveys of heavy metals in aquatic organisms (by Pasminco) since 1972 at 25 sites. There has been bimonthly monitoring of physical parameters, nutrients, chlorophyll-*a*, algae (by E&P/DELM), during 1994–98 at 20–25 sites.

Nutrient concentrations in the Derwent Estuary were monitored over a one-year period, commencing in March 1993. Water samples were collected fortnightly at 51 stations, both at the surface (0.1 m) and at depth (5–20 m). This program was extended in 1996 to 36 sites for nutrients and for various other parameters such as temperature, salinity and dissolved oxygen concentrations at approximately six-week intervals over two days, for a total of nine surveys in 1996.

In 1999, the Derwent Estuary Program (DEP) was established — a joint state, local government and Commonwealth initiative to restore and protect the Derwent Estuary. The primary objective of the first two years of the DEP is to develop a coordinated environmental management plan for the Derwent Estuary, together with agreements for the implementation of specific projects. The intention is to integrate, streamline and standardise monitoring activities in the Derwent Estuary, and to establish a regular reporting and review process. It is also intended that the DEP will draft Protected Environmental Values for the Derwent Estuary.

Table A4.3. List of measurement parameters

Measurement Parameters		Industrial & municipal discharges	Derwent River	Estuary water quality	Present studies (1999–)
Physical	Turbidity		✓	✓	✓
	Colour		✓	✓	
	Temperature	✓	✓	✓	✓
	Total suspended solids	✓	✓	✓	✓
	Electrical conductivity	✓	✓	✓	✓
Chemical	pH	✓	✓	✓	✓
	Dissolved oxygen	✓	✓	✓	✓
	Biological oxygen demand	✓		✓	
	Chemical oxygen demand	✓		✓	
	Ammonia	✓	✓	✓	
	Nitrate + nitrite	✓	✓	✓	✓
	Total nitrogen	✓	✓	✓	✓
	Orthophosphate	✓	✓	✓	✓
	Total phosphorus	✓	✓	✓	✓
	Zinc	✓	✓	✓	✓
	Cadmium	✓		✓	✓
	Copper	✓		✓	✓
	Lead	✓		✓	✓
	Iron	✓	✓	✓	
	Manganese	✓	✓	✓	
	Mercury	✓		✓	✓
	Selenium	✓		✓	
	Arsenic	✓		✓	
	Fluoride	✓		✓	
	Sulfate	✓		✓	
Others	Oil & grease	✓		✓	
	Resin acids	✓		✓	
	Phenols	✓		✓	
	Polycyclic aromatic hydrocarbons (PAHs)	✓		✓	
	Hydrocarbons	✓		✓	
	Chlorophyll-*a*			✓	✓
Biological	Faecal bacteria	✓	✓	✓	✓
	Algal biomass & composition			✓	

Currently (1999–2000), there is quarterly monitoring of physical parameters, suspended solids, nutrients, chlorophyll-*a* and heavy metals at 25 sites throughout the estuary. Physical parameters are measured through the water column and water samples are collected at the surface and at depth. Recreational water quality is also monitored weekly during summer at 30 sites, for faecal coliforms and enterococci.

Other recent monitoring initiatives in the Derwent include the first full survey of estuarine habitat — macroalgae, seagrasses and wetlands — carried out by the Tasmanian Aquaculture and Fisheries Institute (University of Tasmania), and an associated surface sediment survey.

A4.3.4. Data Analysis and Interpretation

Analysis has been limited to simple statistical analysis of medians, ranges and trends (time and distance downstream). Financial constraints have limited the use of models in analysis. Data from the present monitoring programs are being entered into the E&P/DELM database.

A4.3.5. Reporting

Reports have been prepared by the Department of Environment and Land Management (Coughanowr 1995), the Inland Fisheries Commission (e.g. Davies et al. 1988, 1989) and Pasminco EZ. Information from these studies has been presented at workshops and conferences, published in scientific journals, and incorporated in reports by Commonwealth agencies. In 1997, all available environmental monitoring data for the Derwent Estuary were reviewed, compiled and synthesised in the *State of the Derwent Estuary Report* (Coughanowr 1997). The report, which can be accessed at www.derwentriver.tas.gov.au, was an important catalyst for improved management, including monitoring. A Stage 2 discussion paper (Existing Situation) has also been prepared to establish: Uses and Values, Protected Environmental Values, Management Systems and Structures, Environmental Conditions and Key Environmental Issues.

A number of Priority Environmental Issues have been identified and briefing sheets have been prepared for each of these issues for consideration in the next stage of the Derwent Estuary Program. These issues are: sewage discharges, industrial discharges, urban runoff, boat wastes, seafood safety, recreational water quality, heavy metal contamination, organic enrichment/low dissolved oxygen, introduced marine pests, estuarine habitat loss, threatened species, foreshore management, nutrients, upper catchment flows. These briefing sheets provide a good example of the use of existing information to identify and develop objectives and monitoring strategies that meet the information requirements for those objectives.

This review has resulted in the re-establishment of a quarterly monitoring program for heavy metals in the Derwent Estuary and a recognition of the need for a better understanding of the sediments in the estuary, particularly the conditions that may stimulate the mobilisation of heavy metals from the sediments, and how the sediments affect dissolved oxygen concentrations. It has also been recognised that there is a need for continuous monitoring of dissolved oxygen; previously the monitoring had been confined to individual readings during daylight hours, when dissolved oxygen levels tend to be higher.

A4.4. Case Study 4: Long-term Chlorophyll Monitoring in the Great Barrier Reef World Heritage Area

A4.4.1. Objectives and Information Requirements

In 1992, the Great Barrier Reef Marine Park Authority (GBRMPA) established a long-term monitoring program of water within the Great Barrier Reef. The major objective of this program was to investigate the long-term trends and regional differences in nutrient status of the waters that comprise the world's largest reef ecosystem.

This investigation relates to the increased sediment loads that have resulted since European settlement, some 200 years ago. In the last 140 years total nutrient input has increased by about 30%. This excess of nutrients has the long-term potential to damage the fragile ecosystem that exists within the Great Barrier Reef.

The information requirements identified at the outset of the study were these:

- What is the current nutrient status and sediment loading of the Great Barrier Reef system?
- What are the likely effects of increased nutrients and sediment on the reef system; are there specific sites within the reef where water quality guidelines are being exceeded?
- What is the nature and extent of excessive chlorophyll concentrations within the reef environment?
- What are the important sources of the nutrients that are contributing to increased concentrations of chlorophyll, and what would be considered an acceptable level?
- What is the current status of eutrophication within the Great Barrier Reef environment?
- Are there relationships between specific catchment activities and excessive chlorophyll concentrations?
- What is the relationship between chlorophyll concentration and phytoplankton biomass?
- What land management practices are being employed at present within the catchments associated with the Great Barrier Reef?
- *Trichodesmium* exist in the Great Barrier Reef; what is their response to nutrient availability?
- What are the temporal and spatial variations associated with river plumes within the reef?

A4.4.2. Study Design

Conceptual Model

These were the initial assumptions or conceptual models interrelating pressure and state conditions:

- that chlorophyll concentrations within the Great Barrier Reef are a function of nutrient input which is directly related to sediment loading;
- that there is a relationship between catchment-based activities and sediment loading output from that catchment;
- that excessive concentrations of chlorophyll are associated with eutrophication problems, which will lead to the breakdown of reef ecosystems;
- that aquatic grazing has no major influence on phytoplankton abundance and distribution.

More recently, some consideration has been given to event-based sedimentation and nutrient inputs into the Great Barrier Reef. Stevens et al. (1996) studied the spatial influence and composition of river plumes in the central Great Barrier Reef. They concluded that flood plumes can introduce large quantities of nutrients and sediment, resulting in levels that far exceed non-flood conditions. It has also been established that plumes of this nature can travel large distances from the river mouth (Brodie & Furnas 1996), and persist as recognisably distinct water masses for several weeks. It was also stated that these river plumes, within inshore fringing reefs, may persist for longer periods and in a more concentrated form with effects that may be both more acute and longer lasting. Brodie (1996) attempted to relate the catchment conditions, based on climatic zonation within the Great Barrier Reef, to plume characteristics. He concluded that although larger dryland catchments have extreme flow patterns compared to the more even discharge patterns of the 'wet tropics', the extent of their river plumes is often contained to near-shore environments. Geostrophic forces and wind appear to be the major determinants of the directions of the plume travel and extent of dispersion.

Study Boundaries

The monitoring program covers an area of approximately 350 000 km^2 with an archipelagic complex of over 3000 reefs. Almost all the catchments that drain into the near-shore environments of the Great Barrier Reef are used for agricultural purposes and have been extensively modified.

Table A4.4. Weather and hydrographic parameters measured at each station

Parameter	Unit	Method of determination
Weather		
Wind speed	kph	Anemometer
Wind direction	360 degrees	Ribbon and compass
Wave height	m	Visual estimation: trough to peak
Wave direction	360 degrees	Compass
Cloud cover	octets (*/8)	Visual estimation
Rainfall	mm	Gauge
Hydrographic		
Depth	m	Vessel depth sounder
Salinity	%	Salinometer or refractometer in field
Temperature	degrees C	Field meter or thermometer
Secchi depth clarity	m	Secchi disc
Trichodesmium	I/O	Visual: present or absent
Chlorophyll-*a*		Fluorimeter

Measurement Parameters

The parameters to be monitored to respond to the identified questions were:

- the mass of sediment discharged from catchments during event-based discharges, expressed as a function of depth of runoff and land use and management practices;
- the mass of nutrients discharged from catchments during event-based discharges, expressed as a function of depth of runoff and land use and management practices;
- the concentrations of chlorophyll-*a* at inner shore areas and outer shore areas, documenting the season in which samples were taken, water temperature, etc. Specific measurements included chlorophyll-*a*, salinity, temperature, phosphorus, nitrate, nitrite, ammonia, dissolved organic carbon and free amino acids together with supporting general water quality parameters.

Table A4.5. Long-term monitoring transect summary

Marine Park	Sampling transect	Date begun	Shelf range	No. of stations
Far Northern	Cape Weymouth	October 1996	Inshore to offshore	5
Far Northern	Cape Melville	October 1996	Inshore to offshore	4
Cairns	Lizard Island	January 1993	Offshore	5
Cairns	Port Douglas	December 1992	Inshore to offshore	5
Cairns	Cairns	December 1992	Inshore to offshore	7
Central	Townsville	October 1996	Inshore to offshore	4
Central	Whitsundays	October 1996	Inshore to offshore	8
Mackay/Capricorn	Keppels	January 1993	Inshore	5
Mackay/Capricorn	Heron Island	January 1993	Offshore	5

Sampling Sites and Frequency

The design of the monitoring sites, sampling frequency and selection of analytes were guided by the following principles:

- the need to monitor concentration and loads at near-shore locations (<20 km) and offshore locations, with sampling areas being determined by the availability of personnel equipped to undertake sampling;

- the need for event-based monitoring in conjunction with routine monthly sampling;
- the need to quantify regional and cross-shelf patterns of phytoplankton biomass within the Great Barrier Reef lagoon which may be related to regional differences in nutrient inputs;
- the need to determine how much temporal variability (seasonal, event-related) in phytoplankton biomass may reflect changing nutrient inputs to Great Barrier Reef waters;
- the need to monitor ambient concentrations of chlorophyll which would represent relative nutrient concentrations, i.e. the use of chlorophyll as a bioindicator of ambient nutrient concentrations;
- the need to monitor salinity because this is a function of the proximity of the river plume;
- the need for Secchi disc measurements that determine water clarity because this also relates to the proximity and intensity of the plume;
- the need for special studies to better describe in-reef physical, chemical and biological processes, such as:
  - phytoplankton in the Great Barrier Reef and their response to nutrient availability,
  - the effect of excessive nutrient input on reef communities: a physical, biological and chemical perspective,
  - an investigation into physical characteristics and nutrient dynamics of the river plumes and correlations between benthic composition and water nutrient concentrations,
  - nutrient spiking to assess the effects of varying concentrations on phytoplankton biomass, assessment of the duration and proximity of effects in relation to plume intensity,
  - an investigation into the dynamics of phytoplankton grazing within the reef.

Sampling at monthly intervals would be sufficient, giving results that represented seasonal variation. The Great Barrier Reef is spatially immense, so the sampling locations were dictated by the proximity of available personnel. The sampling regime was broken into six main components:

- over 100 stations were monitored once or several times per year; these sites corresponded to previous research conducted on the reef in the late 1970s; the stations were arranged in cross-shelf transects that spanned the broad latitudinal range between 14°S and 23°S extending from the coast to the shelf break;
- at monthly intervals, 38 stations lying along 10 cross-reef transects were sampled; hydrographic parameters, chlorophyll and clarity were measured;
- the weekly and monthly sampling of cross-shelf transects took duplicate surface and near bottom samples with statistical parameters being evaluated in each case;
- provisions for event monitoring, for example, during flood plumes, cyclonic resuspension events and *Trichodesmium* blooms, were built into the sample design;
- built into the sampling design was a smaller-scale study located at Port Douglas (16°S). This involved five sites that were monitored at weekly intervals for dissolved organic matter, dissolved free amino acids, nutrients, phytoplankton biomass and composition, and *Trichodesmium* abundance.
- river nutrient and sediment flux monitoring studies were continued, building on previous research done on the reef (by the Australian Institute of Marine Science (AIMS)). Further refinement of nutrient budget models, both in the Great Barrier Reef lagoon (by AIMS) and for catchment processes (by Queensland Department of Primary Industry), were continued.

It was anticipated that there would be problems of high variability of phytoplankton abundance that would result in numerous problems in the statistical analysis. Trend analysis methods that allow for missing values, non-normal distributions, censored data, autocorrelation in space and time and seasonal trends were used. There was some coordination of the program with other monitoring occurring in the area. This involved long-term monitoring of temperature (by GBRMPA), photographic benthic monitoring (by GBRMPA and Queensland Department of Environment and Heritage (QDEH)), inshore and estuarine water quality monitoring (by QDEH) and sediment monitoring (by AIMS).

A4.4.3. Sampling and Analysis

During this study all the data were collected at the site locations. Among other steps taken for quality assurance, the fluorimeter for chlorophyll-*a* analysis was calibrated regularly against diluted chlorophyll extracts prepared from log-phase diatom cultures.

Data Archive

There was early recognition of the importance of establishing an archive of water quality data, with checking and validation protocols for the entry and management of data, and documentation of sampling and analytical techniques and methods. In view of the volume of the data and the range of groups participating in sampling and analysis, only by this means could there be:

- systematic logging, validation and entry of data;
- secure storage;
- consistent nomenclature, procedures and analytical methods;
- ease of data access.

Entries into the data archive comprised information on the sample data, field observations or comments, sampling method, and site descriptions of reef condition at the time of sampling. User-specified reports were based on site description (number), date and determinands required.

Throughout the study the main principles that guided data management were as follows:

- validation checks were standard for various fields in the database to minimise data entry errors;
- data analysis and interpretation were peer reviewed.

A4.4.4. Data Analysis and Interpretation

The concentrations of chlorophyll recorded at the reef stations were often skewed and it was necessary to transform the data. Typically, $\log_{10}(x)$ transformations were applied which conformed with parametric assumptions of normality and heterogeneity. Extreme values make the choice of summary statistics important. The means are greatly influenced by the outliers (extreme values), whereas median values are more conservative but use only 50% of the data. Techniques such as trimmed mean and M statistics have the advantage of retaining most of the information, yet are obviously more robust to the presence of outliers.

The analyses of long-term temporal trends in chlorophyll concentrations predominantly used non-parametric techniques. This was due to the non-normal nature of the data produced when monitoring chlorophyll concentrations; other factors such as extreme values, serial correlation and seasonality contribute to this issue. One technique that can be employed to analyse such data sets is the seasonal Mann–Kendall tau test. This test involves decoupling monotonic trends in water quality parameters from seasonality. The test sums the number of positive differences between an observation and all later observations, minus the sum of all negative differences. The trend slope is estimated from the seasonal Kendall slope estimator, which can be characterised as the median annual change adjusted for seasonality. This estimator is resistant to extreme values and seasonality. All tests are two-sided, since both the upward and downward trends are possible.

A4.4.5. Reporting

Reporting has comprised Project Reports, annual reviews of water quality, and web-based information. Access to the water quality database is available through the Web.

A range of techniques has been used to present the data analyses, including:

- time-based trend lines of water quality;
- box and whisker plots (1 standard error of mean) for a range of sites and determinands;

- analysis of temporal and spatial variation in chlorophyll concentrations (all data being $\log_{10}$ transformed);
- graphs of phytoplankton nitrogen vs. time, in response to a constant load exceeding the threshold load by 20%;
- monthly surface temperature, salinity and chlorophyll graphs from the stations, produced for both inshore and offshore stations;
- summary statistics of regional chlorophyll-*a* concentrations grouped by season (summer and winter) and cross-shelf position (inshore and offshore);
- frequency observations of *Trichodesmium* slicks during sampling events in each cluster;
- summary of paired *t*-tests between near-shore and near-bottom chlorophyll-*a* concentrations within each regional cluster and grouped by season (summer and winter) and cross-shelf position (inshore and offshore);
- mean (SE±) variability of chlorophyll-*a* concentrations from replicate casts from surface waters, and between duplicates derived from the same cast;
- results of the seasonal Mann–Kendall test for trends in regional and cross-shelf clusters.

Appendix 5

Statistical Methods for Water Quality Monitoring Programs

A5.1. Application of Statistical Procedures in Chapter 6

The purpose of this appendix is to provide additional material and worked examples to clarify the application of the statistical procedures discussed in Chapter 6, and to assist with interpretation of statistical output in the context of water quality sampling, monitoring and assessment. The reader seeking more detail and discussion should consult appropriate references cited here and in Chapter 6.

A5.1.1. Summarising Data

Tables 6.2 and 6.3 in Chapter 6 identify a number of common statistical measures for describing the 'average' and 'spread' respectively of a distribution. However, not all summary statistics are appropriate to all types of measurement. The arithmetic mean is most appropriate for data measured on the interval and ratio scales. The median can also be calculated at the ordinal scale, while the mode can be determined at any measurement scale. Table A5.1 lists the most common summary statistics and the measurement levels required for their determination.

Table A5.1. The applicability of various summary techniques to data measured at each of four levels of measurement

Data reduction method	**Data type**			
	Nominal	Ordinal	Interval	Ratio
Frequency tabulations	✓	✓	✓	✓
Bargraphs	✓	✓	✓	✓
Mode	✓	✓	✓	✓
Evenness Index	✓	✓	✓	✓
Median	✗	✓	✓	✓
Quartiles and percentiles	✗	✓	✓	✓
Interquartile range	✗	✗	✓	✓
Histograms	✗	✗	✓	✓
Frequency polygons	✗	✗	✓	✓
Arithmetic mean	✗	✗	✓	✓
Standard deviation	✗	✗	✓	✓
Variance	✗	✗	✓	✓
Coefficient of variation	✗	✗	✗	✓

- Nominal data are measurements assigned to one of several classes. These measurements are simply counts. The order in which the classes are presented is quite arbitrary, e.g. sex (male or female), maturity status (adult or juvenile), fish species assemblage (Carp, Bass, Perch, Murray Cod, or Redfin).

- Ordinal data are measurements with all the properties of nominal data, but these classes can be ranked in some order, e.g. algae at each site can be absent, sparse, common, abundant, or very abundant. We can say one measurement is larger than another, but we cannot say by how much.
- Interval data are measurements with all the properties of ordinal data; they can also be used to determine by how much one measurement differs from another. Values are measured with respect to an arbitrary zero. Temperature measured in degrees Celsius is an example where comparisons relate to an arbitrary zero, the freezing point of water.
- Ratio data are measurements with all the properties of interval data, but the values are measured with respect to a true zero. An example is phosphorus concentration; a zero means there are no phosphorus molecules present.

This classification is due to Stevens (1946).

A number of additional statistical measures are available to describe features other than location and dispersion. The *skewness* coefficient is a measure of a distribution's asymmetry while kurtosis is a measure of 'peakedness' — usually relative to the normal distribution. Symmetrical distributions such as the uniform and normal have zero skewness coefficients. Most statistical software packages have the facility for computing these quantities and formulae are available from statistical texts. Some important theoretical probability models encountered in water quality monitoring are the normal, gamma, and log-normal distributions. The gamma and log-normal distributions are always positively skewed (i.e. their long tail is to the right) — a feature commonly exhibited by water quality data. Examples of skewness and its impact on common measures of location are shown in Figure A5.1(a–c).

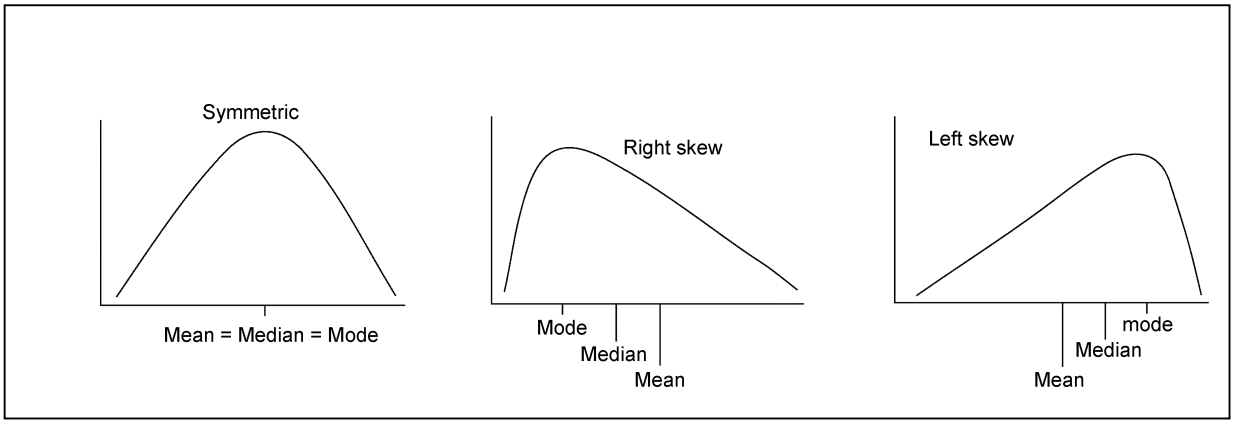


Figure A5.1. Comparison of three distributions and their influence on the measures of central tendency

A histogram of nitrate levels in a water body is shown in Figure A5.2 together with a smoothed estimate of the underlying distribution. The similarity to the gamma distribution (Figure A5.1(b)) and the log-normal (Figure A5.1(c)) distribution is apparent. Histograms are widely used to examine the basic shape and features of a set of data. As Figure A5.2 illustrates, many software packages provide the added facility of using robust smoothing techniques to overlay a 'density estimate' — essentially an estimate of the true or underlying distribution for the entire population of values. Another useful graphical device that permits the extraction of somewhat more quantitative information is the box-plot. The box-plot for the nitrate data is shown in Figure A5.3.

There are a number of ways of constructing and depicting the box-plot, although the essential features are similar. The width of the box typically represents the interquartile range (IQR); the 'whiskers' (lines extending either side of the box) span the range of the data; the horizontal line within the box

depicts the position of the median while the horizontal lines beyond the whiskers denote 'outliers'. Another useful facility offered by most software tools is the ability to generate separate box-plots based on the values of some other factor. For example, the nitrate data depicted in Figure A5.2 are, in reality, data from nine individual sites. Figure A5.4 shows the individual site box-plots. Sites A and B were from one location, sites C, D and H from a second location, and sites E, F, G and I from a third location. These groupings are revealed in Figure A5.4.

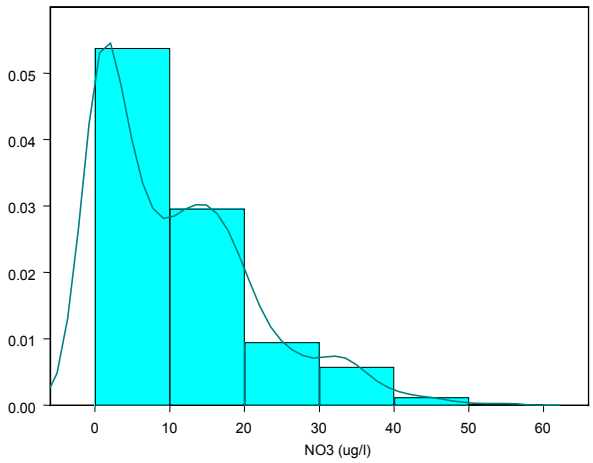


Figure A5.2. Sample histogram of probability densities (y-axis) for nitrate concentrations (µg/L), with smoothed distribution overlaid

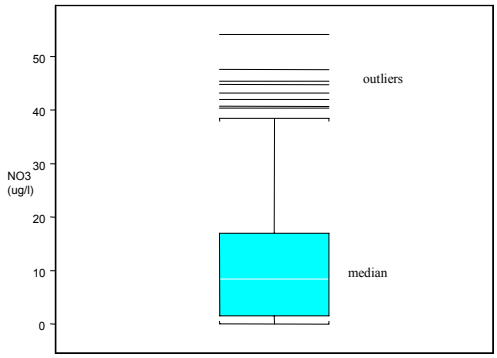


Figure A5.3. Box-plot for nitrate data

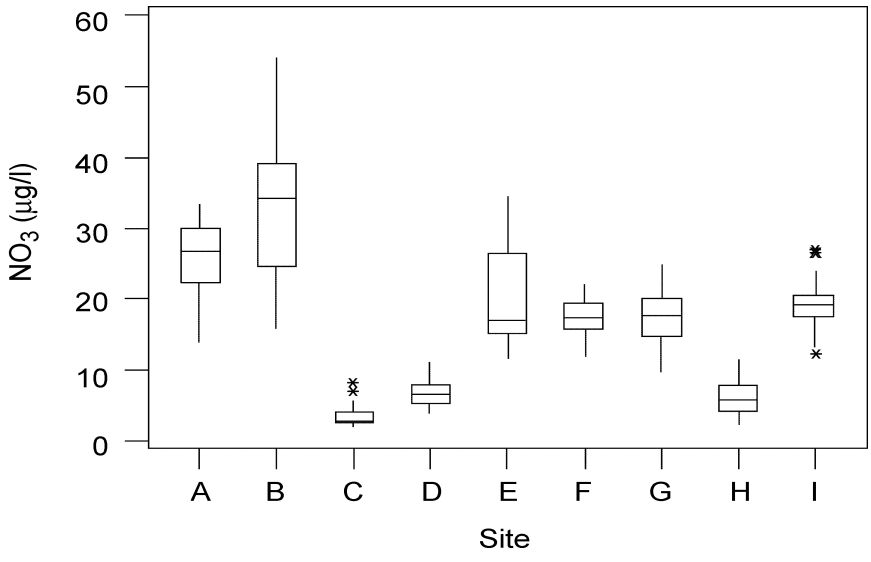


Figure A5.4. Site-specific box-plots for nitrate data (asterisks denote outliers)

Box A5.1. Illustrative panel about summarising data

The nitrate data displayed in Figure A5.2 are investigated in more detail using some standard statistical tools available in most statistical software packages.

Descriptive Statistics

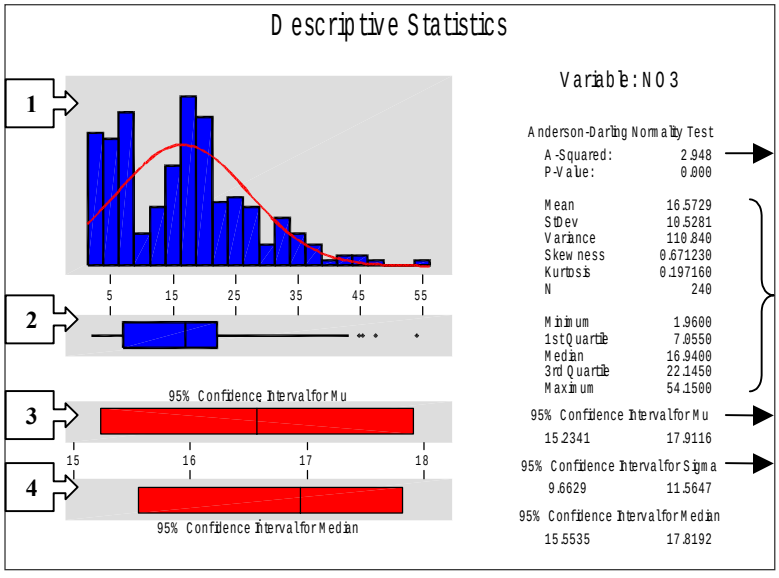
Variable	N	Mean	Median	Tr Mean	StDev	SE Mean
NO3	240	16.573	16.940	15.955	10.528	0.680

Variable	Min	Max	Q1	Q3
NO3	1.960	54.150	7.055	22.145

Explanation:

N — sample size (240 in this case)
Mean — the *arithmetic* mean
Median — 50% of the data have values greater (less) than this value
Tr Mean — a *trimmed* mean obtained by averaging the middle 90% of data (this measure is less susceptible to the influences of extremes and/or outliers in the sample).
StDev —the sample standard deviation
SE Mean — the standard error of the mean. This statistic is a measure of the precision of the arithmetic mean ($SE = \frac{\sigma}{\sqrt{n}}$, where σ is the population StDev).
Min — the minimum value in the data set.
Max — the maximum value in the data set.
Q1 — the first quartile or equivalently, the 25th percentile.
Q3 — the third quartile or equivalently, the 75th percentile.

NB: The inter-quartile range (not shown) is equal to **Q3 – Q1**.



Results of a test of the data's normality. The low *p*-value here suggests the assumption of an underlying normal distribution is untenable.

Descriptive statistics for the data — essentially the same information as presented above.

Numerical information to be read in conjunction with the plot to the left. 95% confidence interval for the *true* standard deviation (see section 6.4.1 and section A5.1.5.1 for discussion of confidence intervals).

1. Histogram of data with smooth estimate of distribution overlaid
2. Box-plot of data
3. 95% confidence interval for the *true* (population) mean
4. 95% confidence interval for the *true* (population) median

A5.1.2. Transformations to Normality

The issue of transforming data prior to statistical analysis is covered in section 6.3.4 in Chapter 6. As noted there, the 'correct' application of many statistical procedures relies on the assumption that the data have been sampled from a parent population of values that are normally distributed. In many instances in water quality sampling, this assumption cannot be made. The Box–Cox transformation is often used in an attempt to restore some semblance of normality to non-normal data. The mathematical form of this transformation is given in section 6.3.4. The identification of the transformation parameter λ is greatly facilitated by a diagnostic plot similar to that shown in Figure A5.5.

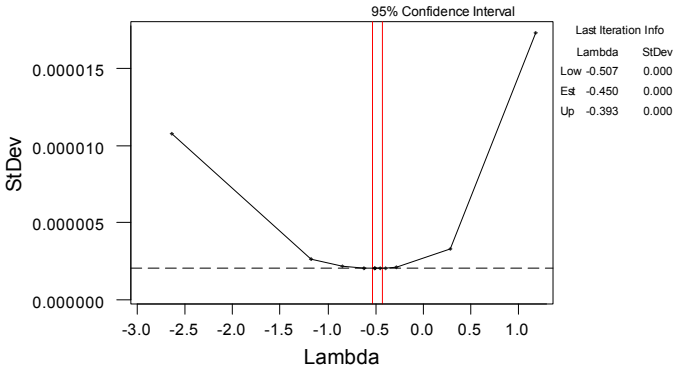


Figure A5.5. Diagnostic plot for the Box–Cox transformation

The optimal value for λ corresponds to the minimum ordinate value in the diagnostic plot of Figure A5.5. The vertical lines define a confidence interval for the true (but unknown) transformation parameter. From the printout on the right-hand side of the plot, we see that the best value for λ is –0.450. Thus, the best power transformation of this data (to improve normality) is

$$Y' = \frac{y^{-0.45} - 1}{-0.45}.$$

Figure A5.6 shows the histogram of the transformed data[1]; the *bimodality* of the transformed data is evident. This is suggestive of two independent processes operating, or else a breakpoint in the data.

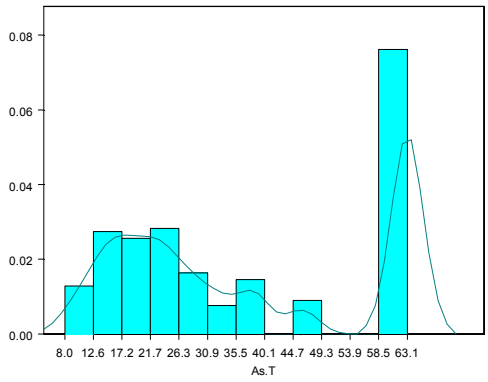


Figure A5.6. Histogram of transformed arsenic data

[1] Note. The software package used in this example has transformed the data using the relationship $Y' = y^{\lambda}$.

A5.1.3. Outlier Detection

The issue of data integrity is covered in section 6.2.2 in Chapter 6. Identification of 'unusual' observations or outliers is reasonably straightforward for single variables. However, as pointed out in section 6.2.2, the task in a multivariate context can be considerably more complex. An example of the joint relationship between two variables is shown in Figure A5.7. The univariate concept of a confidence interval is readily extended to a confidence ellipse in two or more dimensions. The projection of this ellipse onto the axes defines the confidence interval for the respective variable or parameter. The confidence ellipse allows us to make simultaneous inference about all pairs of values (Y_1, Y_2) jointly. For example, we can assert that we are 95% confident that all pairs (Y_1, Y_2) lie within the boundary of the joint confidence ellipse. Thus point P_4 in Figure A5.7 is an 'outlier' in the joint sense, although the one-at-a-time analysis would suggest that this point is not unusual with respect to either Y_1 or Y_2. An analysis of the points in Figure A5.7 is given in Table A5.2.

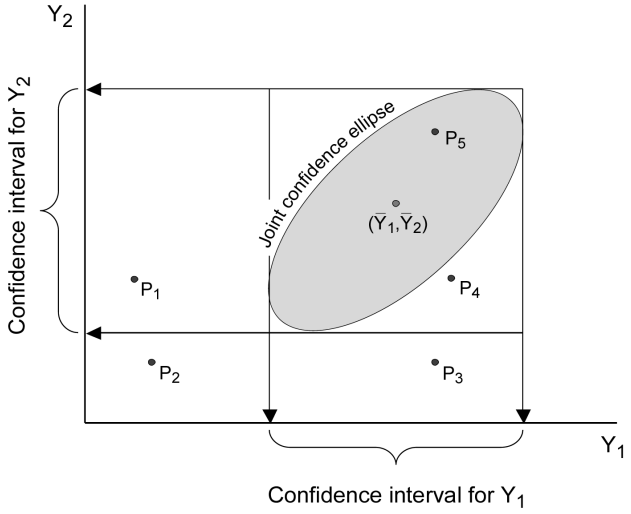


Figure A5.7. Relationship between univariate confidence intervals and the joint confidence ellipse for two variables, Y_1 and Y_2.

The computation of joint confidence ellipses is beyond the scope of this document. The reader interested in learning more about this topic is advised to consult a standard reference on multivariate statistics (e.g. Chatfield 1980).

Table A5.2. Outlier analysis of points in Figure A5.7

Observation	Y_1 outlier?	Y_2 outlier?	Joint {Y_1, Y_1} outlier?
P_1	✓	✗	✓
P_2	✓	✓	✓
P_3	✗	✓	✓
P_4	✗	✗	✓
P_5	✗	✗	✗

A5.1.4. Time Series Analysis

This section on time series analysis complements section 6.3.6 and section 6.6.1 in Chapter 6.

Formal statistical analysis of time series data can be rather complex and requires a good degree of skill in identifying suitable models. Some basic strategies are illustrated with application to the analysis of arsenic in section 6.3.6. Numerous texts have been written on time series analysis, although Cryer (1986) provides a good overview at an intermediate level.

In addition to the assumption of normality, many statistical tests further assume that the data are statistically independent. Violation of this assumption is potentially more serious than violation of the normality assumption and can cause considerable distortion of test results. By their very nature, time series data tend to exhibit varying degrees of serial dependence. A particularly useful tool in identifying and characterising this serial dependency is the autocorrelation function (ACF). The ACF is the correlation between pairs of observations separated by a constant lag or time-step. A plot of the ACF for the arsenic residuals of Figure 6.5b is shown in Figure A5.8.

There are a number of elements to this plot. The vertical lines represent the correlation at successive lags. The envelope represented by the dashed lines is a 95% confidence interval for the true autocorrelation. The simple interpretation is that vertical bars that extend beyond the envelope can be considered to be non-zero, while correlations that are contained within the envelope are indicative of non-significant (i.e. zero) correlations. From Figure A5.8 significant correlations are evident, out to about 15 time lags. The other interesting feature exhibited by Figure A5.8 is the slow linear decay of the ACF. This is suggestive of some structure in the residuals that has not been accounted for by the removal of trend alone. Further insights may be gained from an inspection of the partial ACF or PACF. This is similar to the ACF except that the correlation at lag k, say, is computed in such a way that it is independent of all intervening correlations (i.e. correlations from lags 1 through to k–1). The PACF for the arsenic residuals is shown in Figure A5.9.

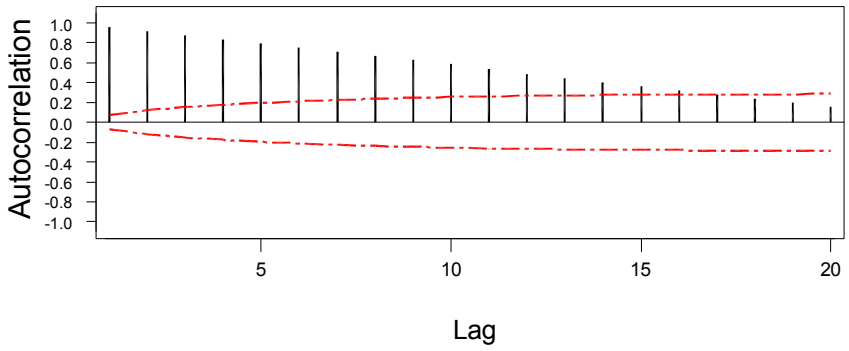


Figure A5.8. Autocorrelation plot of arsenic residuals

From Figure A5.9 the only significant coefficient is at lag 1. The statistician would recognise this as a key feature of a first-order autoregressive model — called AR(1). This simply means that the value of the variable at time (t + 1) is dependent on the value of the variable in the immediately preceding period; that is, at time t. One way of checking this assumption is to 'difference' the series in which a new observation is formed by differencing successive pairs of the original series. Figure A5.10 is a plot of the differenced arsenic residuals, which shows that all structure has effectively been removed and what remains is essentially 'white noise' — that is, random error (albeit with a non-constant variance).

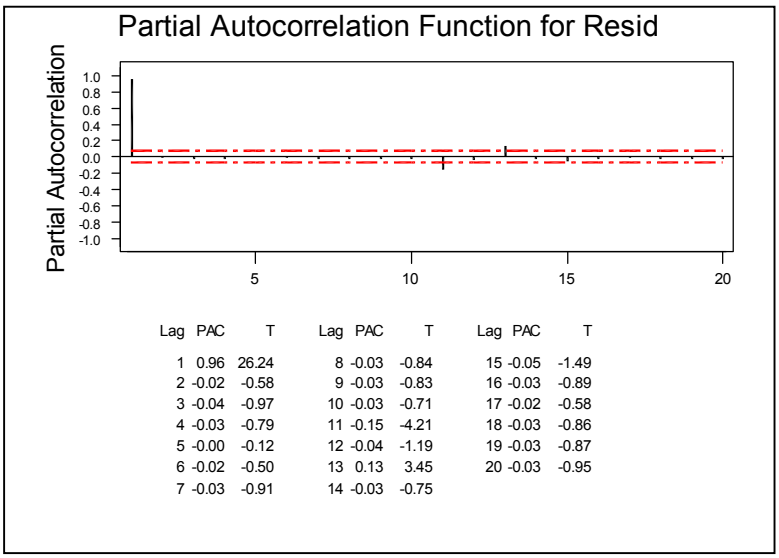


Figure A5.9. Partial autocorrelation function for arsenic residuals

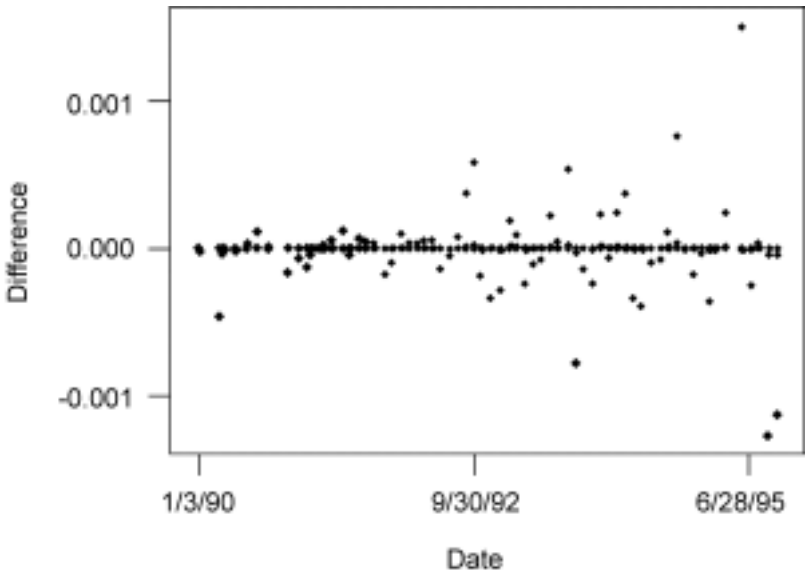


Figure A5.10. Time series plot of first differences of arsenic residuals

A5.1.5. Statistical Inference

The key features of inferential statistics are summarised in Figure A5.11. A population is the largest entity about which some assessment is required (e.g. a lake, a river, a beach, a reservoir, an ocean). Numerical quantities that describe some key feature of the population are referred to as parameters and are designated by Greek symbols. Three common population parameters are illustrated in Figure A5.11.

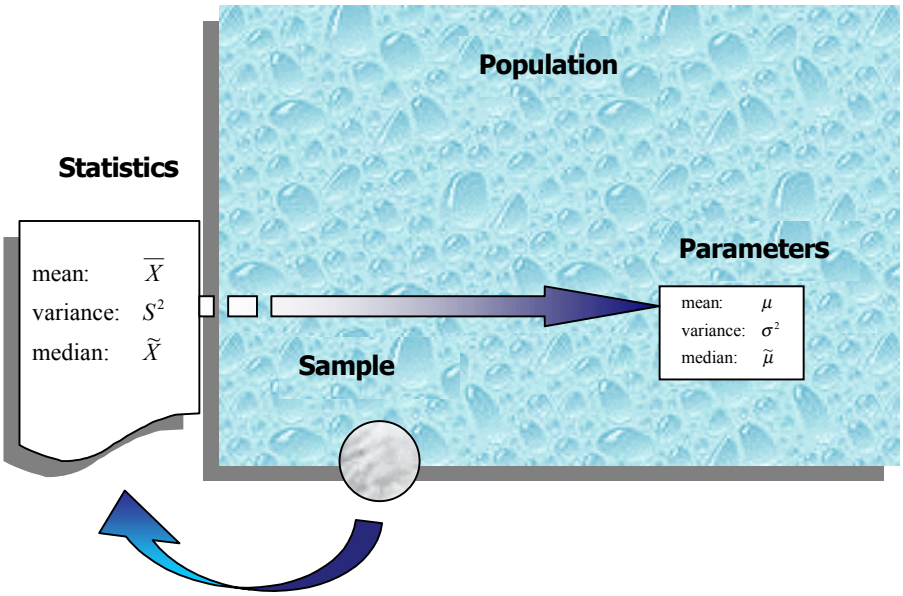


Figure A5.11. Relationship between population parameters and sample statistics

It is usually physically impossible and/or undesirable to sample an entire population and therefore decisions are based on the information contained in a randomly selected sample from the population. The numerical quantities that are computed for the sample are referred to as sample statistics. Sample statistics (or functions of them) are used to make inference about the true (but unknown) population parameters. Methods of inferential statistics fall into two main categories: estimation and hypothesis testing. There is commonality between these two activities although the emphasis is slightly different. Estimation is concerned with assigning a value (or range of values) to the true parameter while in hypothesis testing the aim is to make a judgement about the plausibility of a claimed parameter value. Whereas the reality of sampling is as depicted in Figure A5.11 (that is, a single sample from a larger population of values), methods of statistical inference are established on the notion of repeated sampling from the population. By examining the variation in the values assumed by a sample statistic under repeated sampling, statisticians are able to devise tests that make judgements about the likelihood of a particular result being attributable to chance variation or to some non-sampling induced effect (such as an incorrectly specified hypothesis).

A5.1.5.1. Interval Estimation

The concept of confidence intervals is introduced in section 6.4.1 in Chapter 6. To illustrate the underlying concepts, consider the hypothetical distribution of nitrate concentrations for all water bodies in some region of the country as depicted by the log-normal distribution of Figure A5.12a. The true mean concentration is 14.80 µg/L and the true standard deviation is 10.03 µg/L. The histogram in Figure A5.12(b) represents the distribution of averages obtained by repeatedly taking samples of size $n = 20$ from the distribution in Figure A5.12b. Some important points emerge:

- the population distribution (Figure A5.12a) is decidedly non-normal, yet the histogram of sample means (Figure A5.12b) shows a high degree of normality;
- the 'centre' of the histogram in Figure A5.12b is 14.713, which is very close to the true mean of the parent population, in Figure A5.12a, of 14.80;
- the standard deviation of the histogram of sample means is 2.113, which is less than the standard deviation of the parent population (10.03);
- about 95% of the sample averages in Figure A5.12b are contained in the interval 11 to 19.

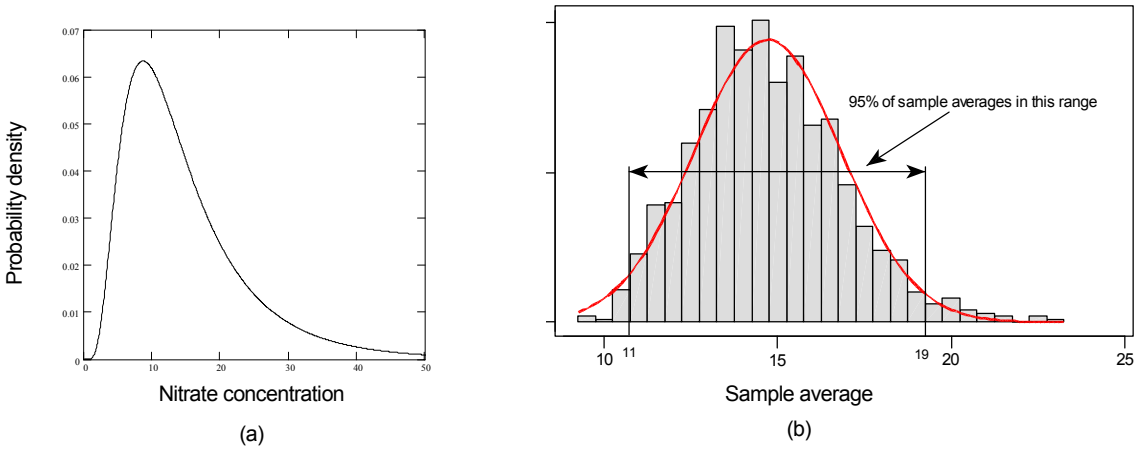


Figure A5.12. (a) Hypothetical distribution of nitrate concentrations for some population; (b) histogram of sample averages based on samples of size n = 20

These four observations are encapsulated in the Central Limit Theorem (CLT) in statistics. The CLT states that:

- the distribution of sample means is normal or approximately so, even if the parent population is non-normal;
- if the mean and standard deviation for the population are μ and σ respectively, then the mean of the distribution of sample averages is also μ, while the standard deviation is $\sigma/\sqrt{n}$ where n is the size of the sample from which the mean was computed. Note, for the example above, σ = 10.032 and so $\sigma/\sqrt{n} = 10.032/\sqrt{20} = 2.24$ which is in close agreement with the standard deviation of the histogram (2.113).

Finally, it is known from elementary statistics that approximately 95% of all observations lie within two standard deviations of the mean. Thus, we would expect that about 95% of our sample means in Figure A5.12b would lie in the interval $\mu \pm 2.0\,(\sigma/\sqrt{n})$; that is, between 10.31 and 19.29 (this is in good agreement with the observed range of 11 to 19). The interval 10.31 to 19.29 is called a 95% confidence interval for the true mean, μ.

To alter our level of confidence, the width of the confidence interval must change — that is the multiplier of 2.0 above needs to change. The appropriate multiplier is found by reference to tables of the standard or unit normal distribution. The general formula for a (1–α)100% confidence interval for a population mean, μ is

$$\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

where (1–α) is the level of significance (as a decimal between 0 and 1) and $z_{\alpha/2}$ is a value from the standard normal distribution (i.e. zero mean and unit variance), so that the area under this normal curve to the right of $z_{\alpha/2}$ is α/2. Table A5.3 gives some common levels of confidence and associated z multipliers.

Table A5.3. Critical z scores for selected levels of confidence

Level of confidence	**90%**	**95%**	**99%**	**99.9%**
z-value	1.645	1.96	2.578	3.291

Confidence intervals provide a good way of comparing observations of a water quality variable to a guideline value, because overlap between the confidence limits and the guideline indicate a lack of evidence that the guideline has not been exceeded.

There is a practical difficulty with the equation given above: namely, it requires knowledge of the true population standard deviation, σ. It is invariably the case that σ will be unknown and must therefore be estimated from the sample. It is tempting to simply replace σ by its sample estimate *s* in the equation above, but this cannot be done without additional modification. Without going into the theory, we need to replace the *z*-value with a value from the *t* distribution. The *t* distribution is very similar to the normal distribution, but it exhibits a degree of spread that is dependent on the sample size *n*. This dependence is expressed by a quantity called the degrees of freedom (df), and for the *t* distribution df = $n - 1$. The effect of various degrees of freedom on the *t* distribution is illustrated in Figure A5.13.

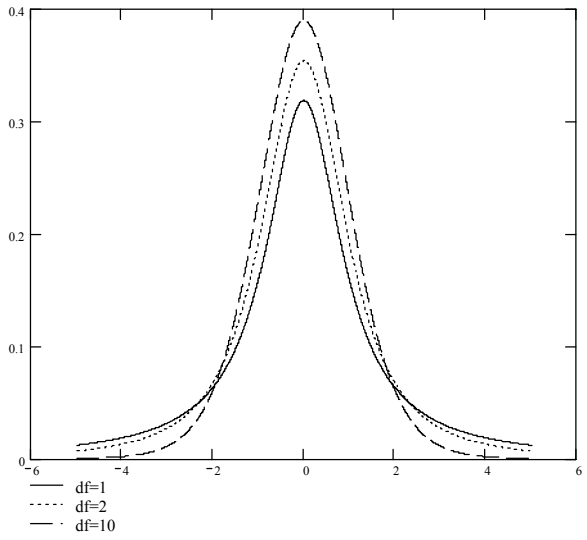


Figure A5.13. *t* distributions for df = 1, 2, and 10

Thus, whenever the sample size is 30 or less the following formula should be used in conjunction with the critical *t* values given in Table 6.6:

$$\overline{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} .$$

For sample sizes greater than 30, the *t* distribution is well approximated by the normal distribution. In such cases the normal distribution can be used to determine the 95% confidence limits.

For non-normal data, it is common to transform the data to yield approximation to normality, then calculate confidence limits for the transformed data. It is possible to obtain approximate limits for the untransformed data by back-transformation.

Alternatively, a confidence interval for the population *median* may be used in place of the population mean. This is the preferred approach when the data exhibit a high degree of non-normality and mathematical transformation cannot increase the normality to an acceptable level.

The steps involved are listed in Table 6.9(a). The data are first arranged in ascending order. The two confidence limits are obtained as the C th observation from each end of the ordered data set. Determination of C depends on the sample size n. When n is 'large' (>20), the formula

$$C = (n - z_{\alpha/2}\sqrt{n})/2$$

is used. Note that C is rounded up to the next highest integer. For small values of n (≤ 20), C is obtained from Table A5.4.

Table A5.4. Values for C for samples of size $n \leq 20$

Sample size	Level of confidence			Sample size	Level of confidence		
n	**0.90**	**0.95**	**0.99**	***n***	**0.90**	**0.95**	**0.99**
4	1	–	–	13	4	3	2
5	1	–	–	14	5	4	3
6	2	1	–	15	5	4	3
7	2	1	–	16	5	4	3
8	3	2	1	17	6	5	4
9	3	2	1	18	6	5	4
10	3	2	1	19	6	4	4
11	3	2	1	20	7	6	5
12	4	3	2				

Example

Determine an approximate 95% confidence interval for the true population median phosphorus level (µg/L) using the following sample data:

4.86, 6.08, 4.83, 4.66, 4.98, 6.16, 6.08, 5.86, 3.95.

The ordered data set is:

3.95, 4.66, 4.83, 4.86, 4.98, 5.86, 6.08, 6.08, 6.16.

Using Table A5.4, $C = 2$ for $n = 9$ at 95% confidence. Thus the approximate 95% confidence limits are the 2nd observations from each end of the ordered data set; that is {4.66, 6.08}.

By way of comparison, a confidence interval for the *mean*, using the formula in section 6.4.1 and a critical t value of 2.306 from Table 6.6 for 8 degrees of freedom ($n - 1$), is

$$5.273 \pm 2.306\left(\frac{0.7916}{\sqrt{9}}\right), \qquad \text{i.e. } \{4.66, 5.88\}.$$

A5.1.5.2. Hypothesis Testing

Several different systems of statistical inference have evolved, and none of them is entirely satisfactory. The major difference between competing systems stems from their attitude to probability. For Bayesian statisticians, probability represents degree of belief in a proposition. This Bayesian view of probability is essentially subjective; probability represents an individual's belief about a system rather than an objective feature of the system. For frequentist or classical statisticians, probability represents the hypothetically limiting relative frequency of an event in a hypothetically infinite repetition of some chance system. For the frequentist statistician, probability represents an objective attribute of the chance system.

The logic of hypothesis testing can be illustrated with a simple example. Suppose the operator of a waste water treatment plant claims that recent upgrades in infrastructure will improve annual compliance with respect to some water quality parameter from the present 70%. The null hypothesis is a minimalist statement that effectively assumes nothing or assumes the status quo. In this example, the null hypothesis (H_0) states that the upgrades have made no difference to the level of compliance. On the basis of sample information, the operator hopes to obtain results that are incontrovertible in their support for the alternative hypothesis — that the true compliance has indeed increased. The parameter to be tested in this example is θ, the probability of compliance on any particular occasion. The situation is reflected by the following pair of hypotheses:

$$H_0: \ \theta = 0.7, \qquad H_1: \ \theta > 0.7 \ .$$

Suppose monthly samples are collected and over the following year the regulator notes that the operator is in compliance on 9 of the 12 occasions, or 75% of the time. Although 75% certainly represents an improvement, the issue to be resolved is whether or not 75% compliance (or better) is likely to occur due to chance even when no improvements have been made. In fact, it can be established that there is about a 50% chance of this happening. This is hardly compelling evidence, and so the decision would be to accept the null hypothesis of no improvement[2]. While this example may seem rather trite, it nevertheless succinctly illustrates a number of key concepts associated with statistical hypothesis testing. First, the analyst is faced with making a binary choice — accept the null hypothesis or reject it[3]. Secondly, this decision is based on incomplete information and thus the decision-making process is subject to error. The types of error possible are identified in Table 6.7 in Chapter 6.

Since the true state of nature is never known, the decision-making process is based on an assessment of probabilities concerning the outcomes identified in Table 6.7. The probability of committing a Type I error is called the level of significance[4] and is denoted by the Greek symbol α. The monitoring team has complete control over α and must specify it in advance of any data analysis. A liberal statistical test will result from assigning a relatively high value for α (since the test procedure will tend to lead to a 'reject' decision more often) whereas setting α to be very small ensures a conservative test. The probability of committing a Type II error (designated by the Greek symbol β) will generally remain unknown because this probability will, in part, depend on the degree to which the null hypothesis has been violated.

A particularly important concept related to β is the complement 1 – β, which is referred to as the *power* of the statistical test. The power of a statistical test is a numerical measure of its ability to correctly reject a false null hypothesis.

In developing hypothesis tests, statisticians first fix the Type I error rate at some small value, typically 0.05 or 0.01. The 'best' decision-rule is the one for which the Type II error rate is as small as possible for this fixed Type I error rate. This treatment of Type I and Type II error rates is asymmetric, in this formulation it is more important to reduce Type I errors.

A5.1.6. The Two-sample *t*-test (Independent Samples)

Often a monitoring team needs to compare two populations. In Worked Example 2 (page A5-29), the analyst wants to know if there is a difference in cadmium concentration between male and female oysters. The two samples have different sample means and sample variances, although this is to be

[2] The probability of getting at least 11 compliances, when in fact the null hypothesis is true, is about 9%; this is still generally not regarded as small enough to assert the veracity of the alternative hypothesis 'beyond reasonable doubt'. Clearly this 'experimental design' appears to disadvantage the operator who would need to demonstrate 12 out 12 compliances in order to 'prove' greater than 70% compliance.

[3] This binary decision making process has been criticised for the narrow focus it forces the analyst to adopt.

[4] The level of significance is also referred to as the *size* of the statistical test.

expected even if the samples have been drawn from the same population. To determine whether or not the observed difference between sample means has occurred simply by chance or whether it reflects a true difference between sexes, a Student's *t*-test can be performed. If the probability (*p*) that the two sample means are from the same population is found to be small (say $p<0.05$) it can be concluded that the two population means are probably different. If the probability is large (say $p>0.05$) the conclusion will be that the difference between means is more likely to be a result of sampling variation.

Before a Student's *t*-test is conducted, the assumption of homogeneity of variances needs to be checked. A number of tests are available for this purpose. If only two variances are being compared, an *F*-test can be used. A generalisation of this test is Hartley's F_{max} test based on the ratio of the largest sample variance to the smallest sample variance; see Ott (1984) for details and a table of critical values. If the variances are found to be not significantly different from one another (see Worked Example 7, page A5-36) a Student's *t*-test may be used to test the means. If the variances are significantly different, a Student's *t*-test would be inappropriate. In this situation Welch's *t*-test, which does not assume homogeneity of variances, could be applied instead.

A5.1.7. The Two-sample *t* -test (Dependent Samples)

The preceding *t*-test required samples to be independent. In cases in which there is a pairwise dependency between samples, the paired *t*-test is preferred: for example, when comparing growth rates of an aquatic plant before and after the administration of a substrate nutrient. A pair of growth measurements is taken from each plant, one measurement before and one after the administration of the nutrient. With paired samples, the procedure is to calculate the difference between members of the pairs, and then to test whether the mean of these differences (as opposed to the difference in means used by the independent samples *t*-test) is significantly different from zero. The paired *t*-test is a more powerful analysis than the independent samples *t*-test because each pair acts as its own 'control', thereby reducing variation due to factors not being investigated. An example of a paired *t*-test is illustrated in Worked Example 3 (page A5-31).

A5.1.8. Analysis of Variance

Analysis of Variance (ANOVA) represents a logical extension of the two-sample *t*-test. ANOVA is a well-established parametric[5] technique for testing the hypothesis of simultaneous equality of a collection of population means. Its major uses in the analysis of water quality data include assessment of the significance of differences in a measurement among water bodies, among different localities in the same water body, or among samples taken from one location at different times.

ANOVA techniques assume that the data in each 'treatment' group are drawn from normally distributed populations, with common variance. The ANOVA procedure is an ingenious device for making inference about the simultaneous equality of the group means by examining components of variance — hence the name. It is possible to obtain two separate estimates of the common variance, σ^2. One estimate, referred to as the within-groups variance, is based on a pooling or 'averaging' of the individual group variances. The second estimate, known as the between-groups variance, uses the fact that the variation between the group means should be roughly σ^2/n. Therefore, if we estimate the variation between the group means and multiply that value by *n*, the size of each group, we obtain an estimate of σ^2. When the null hypothesis is true, these two estimates should be approximately equal, or in other words the ratio of between-group variance to within-group variance should be approximately one. However, when the null hypothesis is false, it can be shown that the between-group estimate is inflated and thus the ratio will return results greater than one. This ratio follows an

[5] In this document the term parametric is used to distinguish statistical methods whose application assumes the sample data follow some prescribed probability model. Statistical procedures that relax this assumption are referred to as nonparametric or 'distribution-free' methods.

F-distribution, and so reference to statistical tables allow us to decide if a particular value of the ratio is too large to ascribe to chance alone when the null hypothesis is true.

Worked Examples 4, 5 and 6 (pages A5-32 to A5-35), are examples of single factor ANOVAs. In Worked Example 4 the analyst wishes to determine if there was a difference in the abundance of mayfly larvae among four shorelines (north, south, east, west). Shore direction is the 'factor', with four discrete 'factor classes', and mayfly abundance is the 'response variable'.

A5.1.8.1. Multiple Comparison Procedures

The rejection of the null hypothesis in an analysis of variance is a rather empty achievement — it enables one to assert with some degree of confidence only that there are differences between the group means. In the example above, the obvious next question is which sites or times are different from which others. To help answer this and similar questions, the monitoring team may employ a multiple comparison technique to help discover underlying causes for the rejection of the null hypothesis.

There are a number of multiple comparison procedures available and they tend to differ with respect to the type of overall error to be controlled for. Procedures for comparing all means with all others include Fisher's least significant difference (LSD), Tukey's multiple comparison test procedure and the Student–Newman–Kuels (SNK) method. Dunnett's test is appropriate where there is a single control site (say upstream of a discharge) and a number of treatment sites (say various distances downstream) with which this is to be compared (see Worked Example 5).

A5.1.8.2. Fixed Versus Random Effects

The above examples are known as fixed single-factor ANOVAs because the factor classes were deliberately chosen. They are fixed in the sense that if the experiment were to be repeated again, the same factor classes would be chosen. This design is the most commonly encountered type in water studies. There is also a random single-factor ANOVA in which the factor classes are randomly chosen from a pool of possible choices. For example, consider the problem of assessing the analytical capabilities of laboratories in measuring chlorophyll-*a*. If there are too many laboratories for all of them to be included, and the monitoring team is not interested in any particular laboratory, a random selection from the pool of possibilities can be made.

A5.1.8.3. Replication and Power

An important consideration in the design of any study is the level of replication to be used. The number of replicates required should be thought about at the design stage of the study. Important factors to consider are: the magnitude of the smallest differences that need to be detected; an acceptable probability that such a difference will not be detected; a notional indication of how variable replicates at a single site and time are likely to be (as indicated by a pilot study); and the significance level of the test to be used. All these considerations are central to a formal power analysis (discussed later in this appendix, page A5-20).

A5.1.8.4. Use of Controls

Another design consideration is the identification of controls. In the context of impact studies (e.g. the effects of an industrial development), a control may be temporal or spatial.

Temporal controls involve the collection of data before an industrial development (say) that may have caused an effect, and form the basis of comparison with post-impact data. The limitation of temporal controls is that there may have been a temporal trend regardless of the industrial development. Attributing a difference between pre- and post-development data to the effect of the development therefore assumes no independent temporal trend. A case must be made in support of this assumption in studies involving only a temporal control.

Spatial controls involve collection of data at a distance from the area likely to have been affected by the development, say upstream from a discharge site, and they form the basis of comparison with the affected sites. The limitation of spatial controls is that the control area and the area subject to potential impact may have been different in the first place. The upstream site may have been different from the downstream sites before the industry was developed and before waste was discharged into the stream. Attributing a difference to the effect of the development assumes there was no prior difference between the control and potentially affected areas. A case must be made in support of this assumption in studies involving only a spatial control.

Other designs have both temporal and spatial controls. Demonstration, before the development, that there is no difference between the area to be developed and the spatial control imparts a certain degree of confidence that any later difference between the affected area and the spatial control area is a consequence of the development. Similarly, demonstration that there is no temporal trend in the control site(s) provides confidence in an interpretation that a difference between pre- and post-development is a direct effect of the development. It is necessary, however, to ensure no spatial–temporal interaction. That is, the monitoring team must be confident that in the absence of the industrial development, the sites that are to be developed would not behave any differently through time compared to the control sites.

A5.1.8.5. Factorial Analysis of Variance

Factorial ANOVA involves the simultaneous analysis of the effects of more than one factor on the measurement variable. Its simplest form, the two-factor design, is more powerful than two separate single factor designs because the effect of one variable can be partitioned out before the effect of the other is tested. For example, in Worked Example 8 (page A5-37), a two-factor ANOVA shows that there was a significant difference in the body mass of *Antechinus stuartii* between the years 1977 and 1978. However, if analysed as a single-factor ANOVA, the difference between years was not significant (Table A5.5). This is because the variation within years was inflated by differences in body mass between sexes (which is demonstrated by the significant result between sexes in the two-factor ANOVA). In the two-factor ANOVA, variability between sexes was partitioned out before years were compared (compare the within-groups sums of squares between the two analyses: the difference is due to sex, 191+333=524). This resulted in a smaller mean square error term and hence a more powerful analysis. Although the researcher may not have been particularly interested in difference between sexes, the inclusion of this factor allowed for a more sensitive test of the temporal effect.

Table A5.5. Results of a single-factor ANOVA on the body mass of *Antechinus stuartii* between the years 1977 and 1978. The data are the same as those used in Worked Example 7 (p. A5-36)

Source	DF	SS	MS	*F*	Pr > *F*
Among groups	1	162.56	162.56	4.34	0.056
Within groups	14	524.38	37.46		
Total	15	686.94			

The two-factor analysis also enables a test for interaction effects. In the above example, body mass may differ between years, but the magnitude of the difference, or, in extreme cases, the direction of the difference, may depend on the sex. In this case the two factors are said to interact.

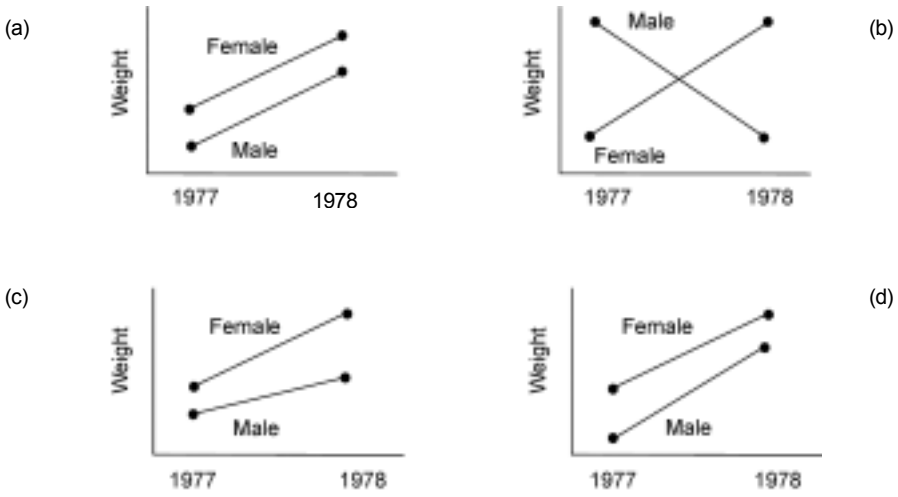


Figure A5.14. Various forms of interactions between year and sex on body mass: (a) represents no interaction; (b) represents strong antagonistic interaction; (c) represents synergistic interaction; (d) represents unimportant interaction

A significant interaction can provide difficulties in the interpretation of the main effects because they must be described in terms of the combined effects of both factors. Figure A5.14b shows a case of an extreme interaction. Differences in mass between sexes are dependent on the year under investigation. Males were heavier in 1977 while females were heavier in 1978. Conversely, differences between years depend on which sex is being investigated. The 1977 samples had heavier males than the 1978 samples, but lighter females. Therefore, in the presence of significant interaction, tests of main effects cannot be made independently of each other. Some interactions (Figure A5.14c) can be corrected by transformation, enabling the main effects to be tested separately. However, a transformation would be of little benefit for interactions such as Figure A5.14b. Some interactions can be considered unimportant (Figure A5.14d) because the effect of one factor on the other is relatively small. In such cases the interaction can be ignored and the main effects tested separately.

A basic strategy in interpreting the results from a factorial ANOVA is to follow these six steps (Neter et al. 1996):

1. examine whether there is an interaction between factors;
2. if there is no interaction, test main effects;
3. if factors interact, determine whether the interaction is important;
4. if interaction is unimportant, proceed as in step 2;
5. if interaction is important, attempt transformation to make it unimportant and return to step 2;
6. if interaction is still important, analyse the two factor effects jointly in terms of the treatment means (e.g. compare body mass between male and female *Antechinus* sp. separately for the years 1977 and 1978, rather than combining the years). See Neter et al. (1996) for more information.

In two- and higher-order factor ANOVAs there are three types of model: fixed factor (all factors are fixed; random factor (all factors are random); and mixed factor (some factors are fixed and some are random). Another important distinction between models is the calculation of the *F* statistic. In a fixed model, the *F* statistic is calculated by dividing the mean square of each factor by the within (error) mean square. In a random model, the *F* statistic is determined by dividing the mean square of both random factors by the interaction mean square, while the interaction is divided by the mean square within. The mixed model is different again. For the calculation of the *F* statistic for each model, see Table A5.6.

Table A5.6. Computation of the *F* statistic for tests of significance in a two-factor ANOVA with replication (Zar 1984); MS is mean square

Effect	Model I (factors A and B fixed)	Model II (factors A and B random)	Model III (factor A fixed, factor B random)
A	MS_A/MS_{within}	MS_A/MS_{AB}	MS_A/MS_{AB}
B	MS_B/MS_{within}	MS_B/MS_{AB}	MS_B/MS_{within}
A B interaction	MS_{AB}/MS_{within}	$MS_{AB}/MS_{wiithin}$	MS_{AB}/MS_{within}

ANOVA allows for the assessment of multi-factor models. Theoretically the number of factors possible is unlimited, but models with five or more factors are rare because of the large number of experimental units that are required and the difficulty in interpreting higher-order interactions.

A5.1.8.6. Nested Analysis of Variance

In the factorial designs discussed above, the factors are crossed, i.e. all levels of one factor occur within all levels of the other factor(s). Both male and female *Antechinus* sp. are included in each year (1977 and 1978). However some experimental designs are not crossed because one factor occurs at different levels in combination with another factor. This is known as a nested design with one or more of the treatments nested within another factor. An example of a nested design can be found in Worked Example 6 (page A5-35), in which fluoride concentration is the dependent variable, and the factor 'sample' is nested within the second factor 'location'. There are three samples within each location, but these samples are unrelated. In this design, the nested factor (sample) is random and the primary factor (location) is fixed. In most cases the nested factor has no intrinsic interest — it is only included to account for some within-group variability. In the example in Worked Example 6, variability due to measurement error has been accounted for by the inclusion of the nested factor (sample). This increases the test's sensitivity to significant effects in the primary factor (location).

A5.1.8.7. Analysis of Covariance

Another important analysis in water quality studies is analysis of covariance (ANCOVA). Suppose a monitoring team wishes to compare the zinc concentration in a species of fish between two lakes (Worked Example 9, page A5-38). A non-significant result is obtained from a single factor ANOVA or *t*-test analysis (Table A5.7). This is because there is a strong negative relationship between zinc concentration and body mass in fish (see graph in Worked Example 9). Therefore, as a large range of fish sizes has been sampled, there is increased variability of zinc concentration in fish within lakes.

Table A5.7. Results of a single factor ANOVA on the zinc concentration in fish, between Lake Arthur and Lake Bull. Data are the same as those used in Worked Example 9.

Source	DF	SS	MS	*F*	*p*
Among	1	2408.33	2408.33	4.38	0.063
Within	10	5493.33	549.33		
Total	11	7901.67			

One way to account for this would be to impose a 'physical control' by using fish of the same size, thereby eliminating any 'size-induced' variation. However this could be time consuming as well as wasteful of data. An alternative strategy is to use a statistical adjustment for the effects of fish size. ANCOVA achieves this by testing for differences among group means after adjusting for group differences in the independent variable (known as the covariate). That is, it compares zinc concentrations in fish between lakes after adjusting for differences in the mass of fish. From Worked

Example 9, it can be seen that there is a significant difference in zinc concentration in fish between Lake Arthur and Lake Bull. Therefore, a large proportion of variability within each lake has been accounted for by the relationship between zinc concentration and body size (compare the within mean square between the two analyses: 11.18 for the ANCOVA and 549.33 for the ANOVA).

Where there are more than two treatments, multiple comparison tests can also be applied to determine which treatment differs from which, although it is important that treatment means are adjusted for the effect of the covariate first.

It should be noted that ANCOVA is different from a two-factor ANOVA, as the covariate (in this case fish mass) is a continuous variable, not a factor. In a two-factor ANOVA both independent variables are discrete. More complex models are possible. For more information on ANCOVA see Sokal and Rohlf (1995) or Neter et al. (1996).

A5.1.9 Generalised Linear Models

Although extremely powerful when attendant assumptions are satisfied, the classical methods of inference presented so far have limitations that restrict their more widespread applicability. For example, the use of ANOVA models to make inferences about the abundances of a rare species is likely to yield misleading results because of the inappropriate statistical model conferred by the ANOVA assumptions. In this instance, the response variable (observed counts) is assumed to follow a normal distribution. The use of a continuous probability model to describe an inherently discrete process may not be an issue if the mean of the counts is sufficiently large for the central limit theorem to apply. However, by definition, our rare species will be observed in very low numbers — perhaps with a mean of five or less and the errors introduced by the normal approximation to the underlying discrete distribution may be substantial. No amount of data transformation is likely to remedy the situation and a more appropriate statistical model should be sought rather than trying to coerce our data into a statistical straightjacket that does not fit. Another limitation of classical methods is the difficulty in modelling non-linear relationships between the response variable and the mean. It was these and other considerations that motivated Nelder and Wedderburn in 1972 to develop a more general class of statistical models known as Generalised Linear Models. There is a subtle distinction to be made here between this new class of models and the classical methods of inference that fall under the umbrella of *general* linear models. The generalisation brought about by Nelder and Wedderburn's development is a consequence of the following features:

- explicit handling of non-normal error distributions;
- easy accommodation of non-linear relationships between the response variable and the mean;
- ability to model both qualitative and quantitative data;
- ability to model processes that have intrinsic mean–variance relationships.

While the generalised linear model framework has led to an enrichment of statistical modelling capabilities, the uptake of these tools has been somewhat disappointing. This may in part be attributed to the more advanced statistical concepts, the introduction of new terminology, and a slightly different approach to analysis and interpretation. Further compounding this situation has been the lack of easy-to-use statistical software for fitting generalised linear models, although programs such as SAS® and S-PLUS® have this capacity.

The underlying theory of generalised linear models is beyond the scope of this document and requires solid foundations in linear algebra and calculus (see also Worked Example 10, page A5-40). The original paper by Nelder and Wedderburn is unlikely to be of benefit to readers of this document. More accessible accounts of generalised linear models and their applications may be found in McCullagh and Nelder (1983) and the text by Dobson (1990).

A5.1.10. Power Analysis and Sample Size Determination[6]

The main purpose of power calculations is to assist in the planning and design of monitoring programs. Before large resources are committed to monitoring, the monitoring team would like to be confident that the program is capable of detecting an effect that is considered large enough to be important, with a reasonably high probability. If a monitoring program has low power, it means that even effects large enough to be of interest are unlikely to produce statistically significant results. In this case, the analyst should consider the possibility of increasing the power (usually by using more replicates), or perhaps question whether the monitoring should be undertaken at all.

The power of a significance test is a concept rooted in the Neyman–Pearson view of hypothesis testing. Power is simply the complement of the Type II error rate: that is, power is the probability of *not* making a Type II error (i.e. of not having a false sense of security), given that the null hypothesis is false. A test with high power (at some specified effect size) is therefore very likely to detect effects of the given size.

In deciding on an appropriate level of replication, account must be taken of both Type I and Type II errors. Neyman and Pearson argued that the costs of Type I and Type II errors could not be defined formally, and that the decision procedure should be based on an informal balancing of Type I and Type II error rates. A Type I error rate of 5% has become conventional, and a Type II error rate of 20% (power 80%) is often considered acceptable. Of course, both of these error rates should be determined in advance and set in the context of the problem at hand. No statistical textbook explains how this ideal state is to be achieved.

Power calculations have also been used to assist in the interpretation of results that are not significantly different. Here their use is retrospective ('how do I make sense of the monitoring I performed?') rather than prospective ('how do I undertake monitoring that will be useful?').

There seems to be some uncertainty about how retrospective power calculations should be performed. Some practitioners perform retrospective power calculations at the effect sizes observed in the experiment. Arguably, there is little merit in this approach. Since retrospective power calculations tend to be computed only when the null hypothesis is accepted, the calculated power is likely to be very low. Indeed for a test statistic that has a symmetric distribution under the alternative hypothesis, the power calculated in this way will always be less than 50%. In the extreme case of no observed effect, the power will be equal to the significance level. What is important is the power of the experiment to detect results that are believed to be important. It is recommended that power calculations be performed for a range of effect sizes, spanning the region in which effects are large enough to be important.

The following section illustrates some of the basic concepts behind power and sample size calculations. Given the complexity of the calculations involved, no attempt is made to provide a comprehensive catalogue of formulae for various statistical test procedures — these are best left to reliable software tools. The calculations used in this section have been performed using CSIRO's free software package *PowerPlant*®. A copy of the software and documentation may be downloaded from the following site: ftp://ftp.per.its.csiro.au/csiro-wa/biometrics. A list of other power and sample analysis software tools can be found at: http://www.insp.mx/dinf/stat_list.html or http://www.forestry.ubc.ca/conservation/power/index.html .

A comprehensive review of a number of these utilities can be found in Thomas and Krebs (1997).

A5.1.10.1. Basic Concepts

Consider the following hypothesis-testing situation: H_0: $\mu = \mu_0$, H_1: $\mu > \mu_0$,

[6] Adapted from Statistical Power Analysis – Course Notes, CSIRO Mathematical and Information Sciences www.cmis.csiro.au/envir/Training/StatPowerAnal.htm .

where μ is the true but unknown mean concentration of some water quality parameter, and μ_0 is the hypothesised value. A graphical representation of the distribution of the parameter of interest under each of these hypotheses is shown in Figure A5.15 (note it is necessary to assume normality and a common variance σ^2).

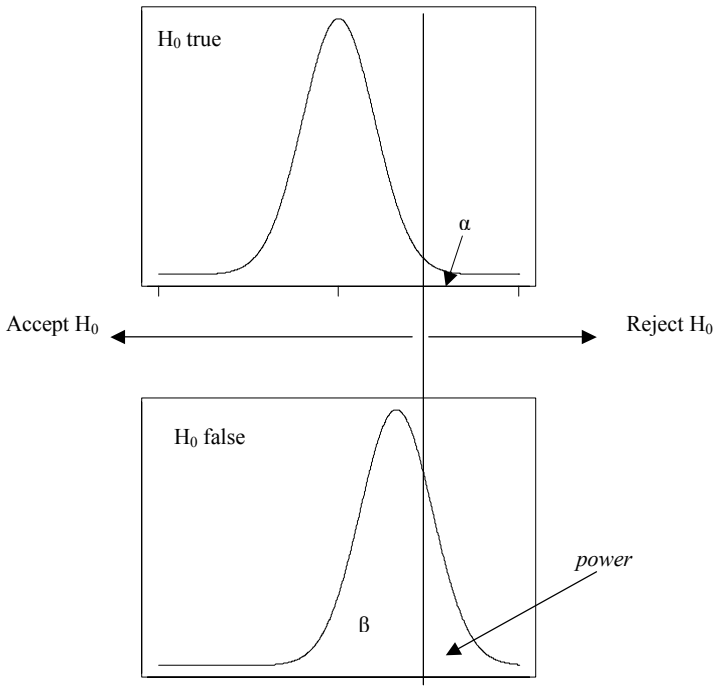


Figure A5.15. Level of significance (α), Type II error (β), and power (1 – β)

When the null hypothesis is true (top curve in Figure A5.15), the probability of incorrectly rejecting H_0 (α) is represented by the area to the right of the decision rule (the vertical line in Figure A5.15). Conversely, when the null hypothesis is false (bottom curve in Figure A5.15), the probability of incorrectly accepting H_0 (β) is represented by the area to the left of the decision rule (the vertical line in Figure A5.15). The power (1 – β) is the probability of correctly rejecting a false H_0 and this is depicted as the area to the right of the decision rule under the bottom curve of Figure A5.15. This simple example serves to highlight the fact that the power depends on the degree to which the bottom curve of Figure A5.15 has been displaced relative to the top curve (i.e. the bigger the 'effect size', the greater the power of the statistical test). Power is also affected by sample size, level of significance, and variance of data.

Example

Consider the situation in which the concentration of some analyte is to be compared at five locations using a one-way ANOVA design. The null and alternative hypotheses are:

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ versus H_1: at least two means different.

The monitoring team statistician has decided on using $n = 4$ samples at each of the five sites. From past experience she knows that the standard deviation is approximately 0.14 µg/L. It is important that the experimental design has sufficient power to detect a minimum difference of 0.02 µg/L using a 0.05 level of significance. The dialogue box from the *PowerPlant*® software is shown in Figure A5.16.

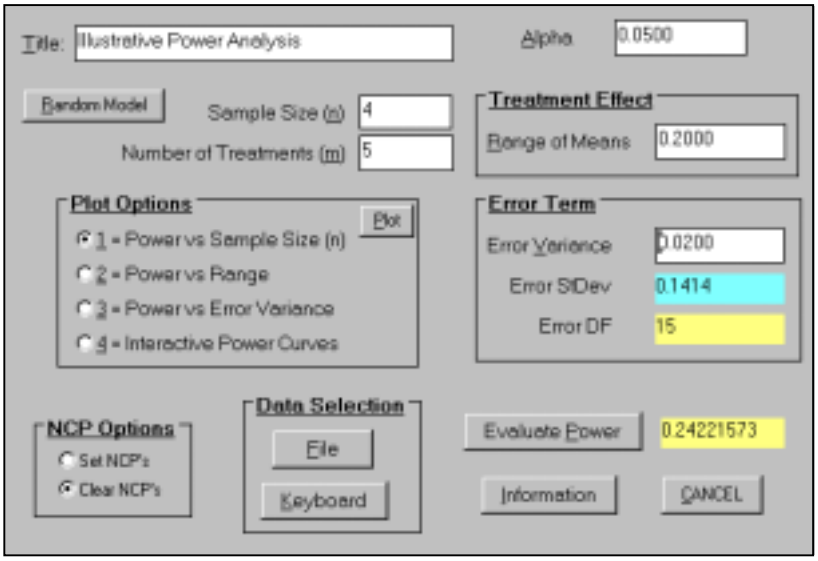


Figure A5.16. *PowerPlant*® dialogue box for power and sample size calculations

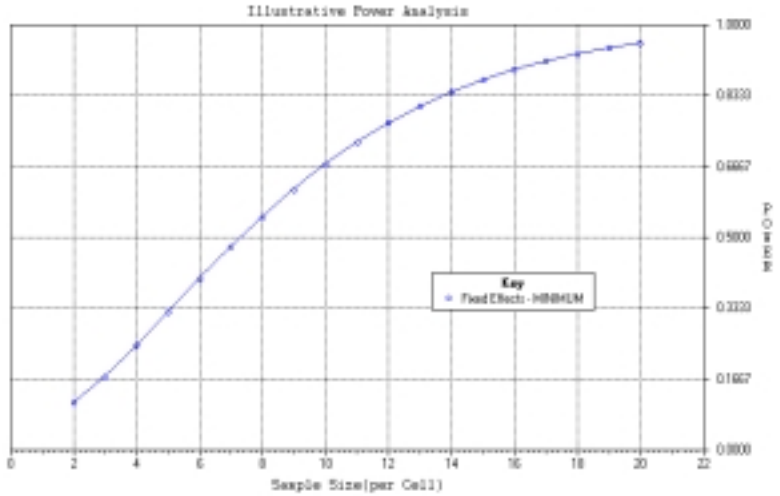


Figure A5.17. Power curve from *PowerPlant*® software

It can be seen from Figure A5.16 that the power of the current experimental design is a little under 25%. This would be regarded as too low to be of benefit and the decision must be made to either abandon the experiment or increase resources (more samples). To help address the latter, it is useful to obtain a plot of the power curve as a function of sample size. This is simply a matter of selecting the appropriate plot option in the *PowerPlant*® dialog box and pressing the 'Plot' button. The result is displayed in Figure A5.17. Figure A5.17 shows that approximately 13 samples at each site need to be taken in order to have a power of approximately 80%. (See also Worked Example 11, page A5-41.)

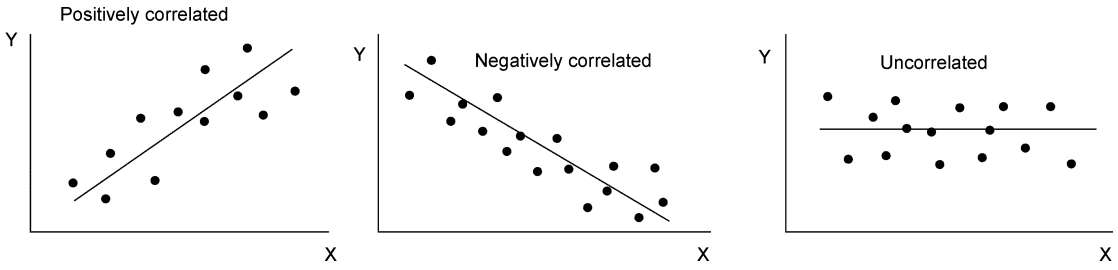


Figure A5.18. Examples of positively correlated, negatively correlated and uncorrelated data

A5.1.11. Correlation and Regression

Correlation analysis refers to the statistical methods associated with quantifying the (linear) relationship between two or more variables. Examples of the types of relationships observed are shown in Figure A5.18. Regression analysis refers to statistical procedures that attempt to *model* or describe (in the form of mathematical functions) that relationship. Both correlation and regression are discussed in section 6.5 in Chapter 6. Regression is illustrated in Worked Example 12, page A5-42.

For simple linear regression, the following model is assumed,

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \ ,$$

where β_0 is the true (but unknown) intercept, β_1 the true slope and ε_i the ith residual. The latter terms are introduced to account for the imperfect fit between Y and x. Furthermore, these residuals are assumed to follow a normal distribution having mean zero and variance σ_ε^2. The regression analysis usually commences (and unfortunately, all too often, stops) with the estimation of the parameters β_0 and β_1. The standard statistical procedure for accomplishing this is known as ordinary least squares (or OLS). Formulas for estimating β_0 and β_1 from sample data are provided below (a 'hat' over a β differentiates an estimate of a parameter value from the parameter value itself):

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \quad \text{and} \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

where

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{1}{n}\left(\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i\right); \quad S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2 .$$

Regression models are used in a variety of contexts. For example, a monitoring team might wish to develop an empirical model to predict chlorophyll-*a* concentrations from loads of phosphorus. Regression can also be used as an exploratory tool to investigate relationships between a response variable and other environmental variables with a view to formulating hypotheses for subsequent manipulation and experimentation to establish causation.

Aquatic systems are often too complex to be modelled by simple linear regression. For example, chlorophyll-*a* concentrations would be dependent on more than just the loads of phosphorus alone. Other variables such as flow rate, water temperature and nitrogen concentration may also be important in determining the concentration of chlorophyll-*a*. Therefore, a simple linear model is likely to be a poor predictor of variation in chlorophyll-*a* concentrations. In such cases a multiple regression model may be more appropriate. The multiple regression model takes the form

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi} + \varepsilon_i \ .$$

The interpretation of the βs is that they represent the change in Y per unit change in the corresponding X, assuming all other Xs are held constant.

An important assumption concerning the error terms of these regression models is that they are independent. Samples collected serially in time often display a degree of autocorrelation. For example, if a measurement taken at one time is above the value predicted from the statistical model under consideration (in this case a regression line), it is likely that the next value will also be above its predicted value. For example, the concentration of phosphorus in storage at a particular time has a great bearing on the concentration an hour later, and probably a day later. If one measurement is well above the general trend, the other is likely to be also. Failure to ensure independence among measurements taken through time can have profound effects on the assumed Type I error rate, though the estimated parameters of the regression remain unbiased.

One way to overcome temporal dependence is to select a sampling interval that is large enough to ensure no connection between consecutive measurements. Alternatively, various autoregressive models are available for analysing time series data, and the reader is referred to the text *Applied Linear Statistical Models* by Neter et al. (1996) for an introduction.

A5.1.11.1. Robust Regression

Robust regression is discussed in section 6.5.3 in Chapter 6. To illustrate some of the concepts, consider the metal data of Figure 6.2. Inspection of Figure 6.2 shows that although the relationship between lead and copper concentrations exhibits a high degree of linearity, there are two lead values which appear to be atypical. The Cu – Pb scatter plot is shown in Figure A5.19 together with the regression line estimated using OLS. The influence of the two aberrant lead values is most pronounced and while the fitted line is the 'best' in terms of the least-squares criterion, its predictive capability would be low. To overcome this difficulty we could remove the offending observations and re-fit the line. This may be appropriate in this instance since the source of the problem is clear. However, in some instances it may not be obvious which observations have high 'leverage' and so a robust method using all the data would be preferred. The robust regression analysis of the Cu – Pb data is shown in Figure A5.20. The resulting line provides a much more realistic fit to the data.

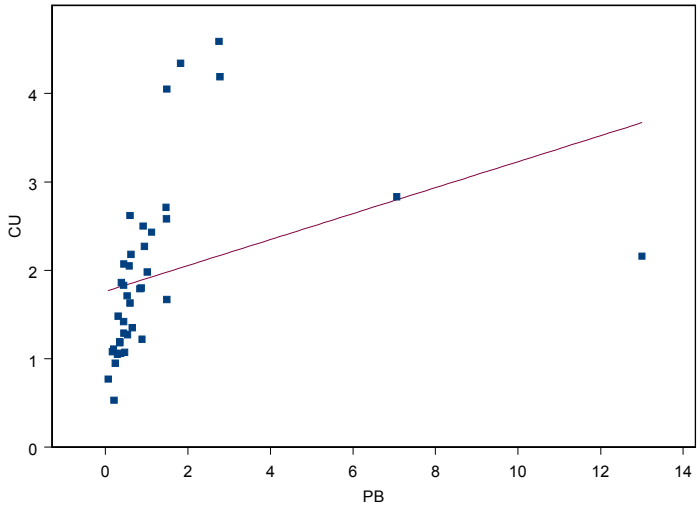


Figure A5.19. Scatterplot for copper and lead concentration data of Figure 6.2 with OLS regression line overlaid

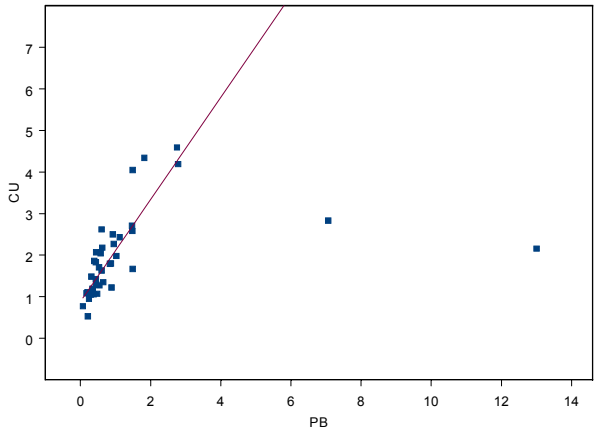


Figure A5.20. Bi-plot for copper and lead concentration data of Figure 6.2 with robust regression line overlaid

A5.1.12. Generalised Additive Models

In the same spirit as the generalised linear models described earlier, generalised additive models (GAMs) have been devised to increase the flexibility of statistical modelling. As noted by Hastie and Tibshirani (1990), 'we can now augment the linear model with new methods that assume less and therefore potentially discover more'. The GAMs represent an extension of conventional regression modelling techniques. Rather than imposing and estimating some predefined model, GAMs replace the usual linear function of an independent variable with an unspecified smooth function. In this sense, the model is nonparametric because a parametric form is not imposed on the functions — they are suggested by the data.

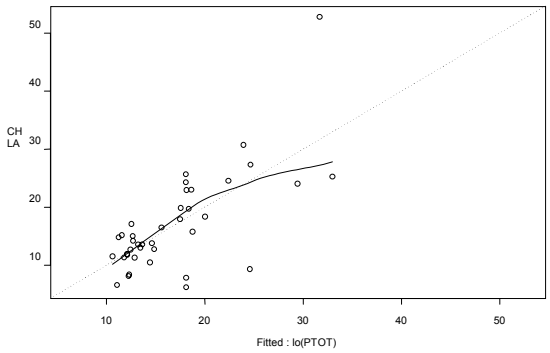


Figure A5.21. Plot of measured chlorophyll-*a* and fitted values using a smooth function of phosphorus

By way of example, consider modelling the relationship between chlorophyll-*a* and total phosphorus in a lake using a generalised additive model. Rather than deciding on some functional form in advance and then estimating the parameters of our model, we can use a GAM to help determine the nature of this relationship. Figure A5.21 shows the result after fitting chlorophyll to a smooth function of total P. The dashed line in the plot is used to gauge the degree of departure from a linear model. From Figure A5.21 we see the departure is quite small, although some non-linearity is evident at higher chlorophyll levels. In order to assess the adequacy of the fit a comprehensive statistical analysis of the resulting model is possible, but this requires an understanding of more advanced concepts. These are covered in the text by Hastie and Tibshirani (1990).

A5.1.13. Nonparametric Statistics

Parametric tests such as *t*-tests and ANOVA have several restrictive assumptions, such as homogeneity of variances and data that are normally distributed. Violations of these assumptions can result in seriously flawed decision-making. In situations where the assumptions of a parametric test cannot be met, non-parametric tests or 'distribution-free' counterparts may be more appropriate. Some of the more common parametric tests and their nonparametric alternatives are listed in Table A5.8. For more detailed discussion on nonparametric tests see Neave (1988).

Table A5.8. Common parametric tests and their nonparametric alternatives, and the advantages and disadvantages of nonparametric tests

Parametric test	Nonparametric test
Paired *t*-test	Wilcoxon matched pairs signed ranks test
Students *t*-test	Mann–Whitney U
	Kolmogorov–Smirnov two-sample test
One-way ANOVA	Kruskal–Wallis H test
Two-way ANOVA	Friedman two-way ANOVA for ranks
	Scheirer–Ray–Hare extension of the Kruskal–Wallis test
Linear regression	Kendall's robust line-fit method
Pearson's correlation *r*	Spearman's rank correlation
	Kendall's rank correlation

Advantages of nonparametric tests	Disadvantages of nonparametric tests
• Free of most distributional assumptions	• Generally less powerful than parametric tests
• Resistant to outliers (this may also be seen as a disadvantage)	• Do not always make most use of information contained in data
• Generally easy to perform	• There are no non-parametric alternatives to some of the more complex parametric analyses such as multiple regression and multi-factorial ANOVA
• Potentially more robust in the presence of non-detects	
• Handle nominal and ordinal data	

A5.1.14. Multidimensional Scaling

A detailed account of multidimensional scaling (MDS) is beyond the scope of this document and the interested reader is advised to consult one of the numerous texts on multivariate statistics, e.g. Clarke and Warwick (1994). This section provides a very brief overview of the method and its uses, and an illustrative example (see also section 6.6.3 in Chapter 6).

Multidimensional scaling is usually introduced by analogy with the following situation. Given a map showing the locations of major cities it is a straightforward task to construct a table of distances between the cities. However, the reverse problem is considerably more difficult. That is, given a table of inter-city distances, how is the map constructed? It is this latter problem that MDS attempts to resolve. The difficulty is compounded in the natural environment because our input data (the

'distance' matrix) is subject to considerable uncertainty, and, furthermore, the number of dimensions required to adequately represent the observed similarities or distances is not always evident. As noted in section 6.6.3 in Chapter 6, the computations required to undertake an MDS analysis require access to specialised statistical software.

An example of an MDS plot is shown in Figure A5.22. In this case researchers were interested in spatial patterns among fish populations in Victoria's Port Phillip Bay. Measurements on fish biomass for a large number of species were recorded from a variety of sites throughout the Bay. A matrix of 'similarities' between pairs of species–site data was obtained and this was analysed by conventional MDS methods. Figure A5.22 reveals three distinct groupings or clusters of observations that the researchers were able to identify with different habitat classifications (pale grey = shallow; grey = medium; dark grey = deep). In this case, the MDS has achieved its objective of extracting a pattern from data that otherwise would be difficult to visualise in many dimensions.

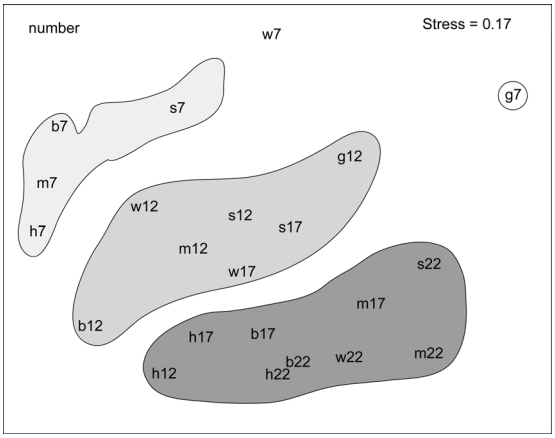


Figure A5.22. MDS plot of fish habitat data in Port Phillip Bay (taken from Parry et al. 1996)

There are differing views about the extent to which MDS can be used as an inferential tool. Computationally intensive methods are available which enable the researcher to conduct formal tests of significance, although some would argue that the strength of MDS lies in its descriptive capability and that it should thus be confined to the realm of exploratory data analysis (EDA).

A5.2. Worked Examples (see next 15 pages)

The worked examples were prepared using MINITAB® statistical software. These examples have been adapted from real analyses. They include actual computer output, represented by typewriter font.

Worked Example 1: Checking Distributional Assumptions

Context

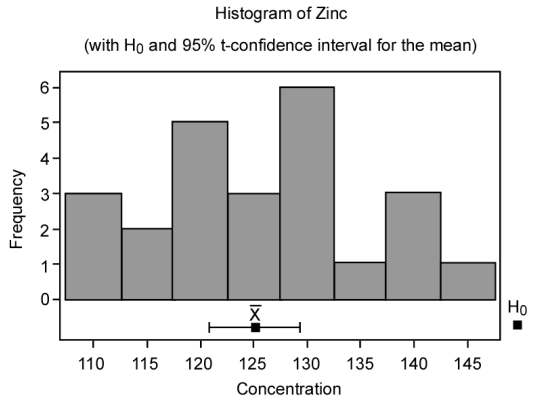
There have been concerns in recent years over the effects that contaminants in our waterways are having on marine life. Organisms such as fish and oysters are known to accumulate heavy metals in their tissues to such dangerously high levels that they can be unsafe to eat. This is especially the case in the Georges River, Sydney. Therefore a research scientist was interested to see if the zinc concentration in fish from the Georges River were at a concentration that would be harmful to the health of humans; 24 fish were caught and the zinc concentrations were determined. The results were as follows:

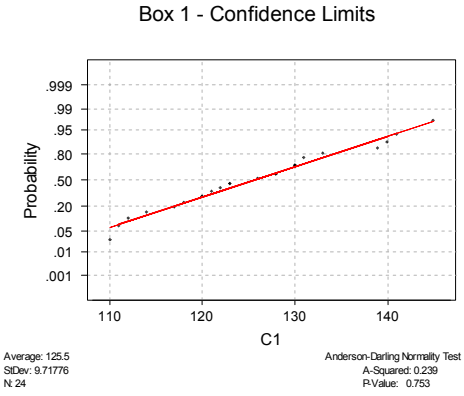
Zn(μg/g)			
141	130	130	112
126	123	133	131
128	139	130	145
128	122	114	117
110	118	111	120
121	140	123	120

Analysis

A useful way of checking on the distributional assumption (normal or otherwise) is via a *probability plot*.

The normal probability plot for the Zn data is shown below. Departures from linearity provide evidence to suggest non-normality. Various formal statistical tests of the hypothesis of normality can also be conducted. Results from the Anderson–Darling test of this hypothesis confirm the impression given by the plot — that the assumption of a normally distributed parent population is tenable in this instance.





The sample histogram is shown above (not overly informative for a small sample size) with an indication of the relative positions of the sample mean, the 95% confidence interval, and the value hypothesised under H_0 (in this case 150 μg/g corresponding to the NFA guideline). In this case, the hypothesised value of 150 is well removed from the extremities of the confidence interval. A formal test of the null hypothesis is also given below:

t-Test of the Mean (see Table 6.9(a) test for a single population mean)

```
Test of mu = 150 vs mu > 150

Variable      95.0% Lower Bound        T      P
Zinc                     122.10   -12.35  1.000
```

Worked Example 2: Two-sample t-test

Context

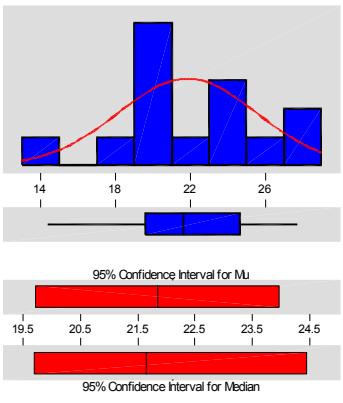
A researcher was interested to know if there was a difference in cadmium concentration between male and female oysters (*Saccostrea crassostrea*). Sample oysters, 13 males and 14 females, were obtained from the Clyde River, Batemans Bay, and analysed for cadmium. The results (in µg/g) were as follows:

Male		Female	
21.4	23.7	27.1	19.7
22.3	21.9	19.4	19.8
28.4	28.2	22.5	27.7
21.7	22.2	25.4	19.9
29.0	17.9	20.8	17.9
24.5	11.7	23.6	14.4
21.4		23.2	24.4

Preliminary Analysis

Descriptive methods (graphical, tabular, and summary) are important tools for teasing out important properties of sample data. The choice of statistical tools has little or nothing to do with the assumed underlying distribution. The panels below show summary presentations of Cd concentrations in male and female oysters.

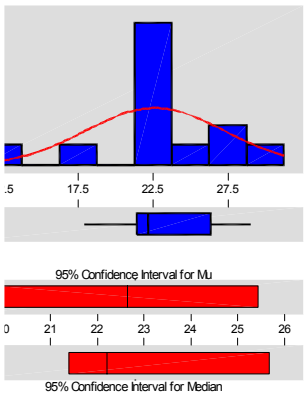
Descriptive Statistics



Variable: Cd - Fem

Anderson-Darling Normality Test	
A-Squared:	0.215
P-Value:	0.811
Mean	21.8429
StDev	3.6903
Variance	13.6180
Skewness	-1.4E-01
Kurtosis	-8.6E-01
N	14
Minimum	14.4000
1st Quartile	19.6250
Median	21.6500
3rd Quartile	24.6500
Maximum	27.7000
95% Confidence Interval for Mu	
19.7122	23.9736
95% Confidence Interval for Sigma	
2.6753	5.9452
95% Confidence Interval for Median	
19.6846	24.4515

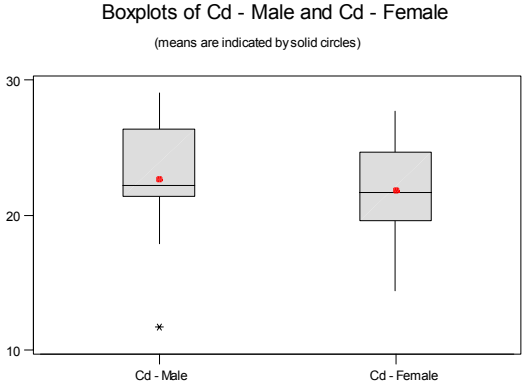
Descriptive Statistics



Variable: Cd - Mal

Anderson-Darling Normality Test	
A-Squared:	0.587
P-Value:	0.101
Mean	22.6385
StDev	4.6196
Variance	21.3409
Skewness	-6.2E-01
Kurtosis	0.130699
N	13
Minimum	11.7000
1st Quartile	21.4000
Median	22.2000
3rd Quartile	26.3500
Maximum	29.0000
95% Confidence Interval for Mu	
19.8468	25.4301
95% Confidence Interval for Sigma	
3.3127	7.6258
95% Confidence Interval for Median	
21.4000	25.6669

Another useful graphical device for summarising and comparing data sets is the box-plot. Sample box-plots for the male and female Cd data are illustrated below. Although the means (and medians) are very similar, the distribution of Cd in the males is (positively) skewed.



Worked Example 2 (continued)

Formal Analysis

Two Sample *t*-Test and Confidence Interval

```
Two sample t for Cd - Male vs Cd – Female
           N      Mean     StDev   SE Mean
Cd - Mal  13     22.64      4.62       1.3
Cd - Fem  14     21.84      3.69      0.99

95% CI for mu Cd - Mal - mu Cd - Fem: ( -2.6,  4.15)
t-Test mu Cd - Mal = mu Cd - Fem (vs not =): t= 0.49  P=0.63  DF=  22
```


```
95% CI for mu Cd - Mal - mu Cd - Fem: ( -2.6,  4.15)
t-Test mu Cd - Mal = mu Cd - Fem (vs not =): t= 0.49  P=0.63  DF=  22
```


Interpretation

The first part of the output above gives the relevant statistics for the difference between males' and females' cadmium concentrations. The means, standard deviations, and standard errors are all in close agreement. The significance of the difference between sample means is tested formally (and equivalently) by reference to the next two lines.

The confidence interval approach shows that a 95% confidence interval for the *difference* (male – female) ranges from –2.6 to 4.15. Since this interval includes zero, we cannot rule out a zero difference. That is, the inequality between male and female cadmium concentrations has *not* been established.

The alternative approach is to conduct the two-sample *t*-test and assess the significance of the result using a *p*-value. We see that the computed *t*-value of 0.49 has an associated *p*-value of 0.63. Since this *p*-value represents the probability due to chance of obtaining a *t*-score of 0.49 (or higher) when the null hypothesis is true, we accept the null hypothesis of equality. To reject this hypothesis at a level of significance α, the computed *p*-value would need to be less than or equal to α.

Worked Example 3: Paired *t*-test for Dependent Samples

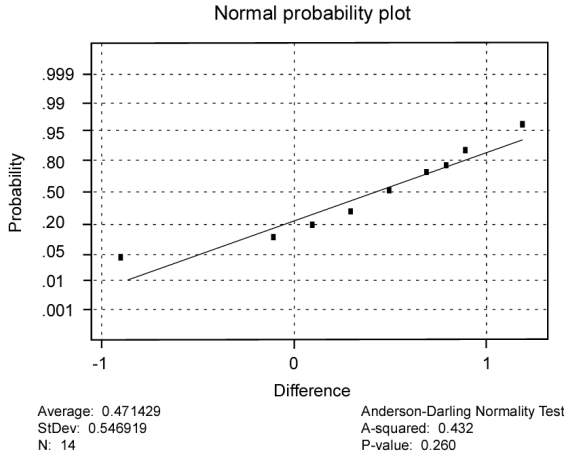
Context

A biologist was interested in determining whether the size of eggs produced by the freshwater turtle *Emydura macquarii* was constant or highly variable between years. Fourteen females from the Macleay River were caught in both 1995 and 1996 and were gravid on each occasion. Females were x-rayed and induced and the eggs were weighed. The average weight of eggs in a clutch for each turtle for the years 1995 and 1996 were as follows:

ID	1995	1996	Difference	ID	1995	1996	Difference
501	7.5	8.3	0.8	536	7.3	6.4	–0.9
504	7.2	7.7	0.5	537	7.6	7.7	0.1
505	6.8	7.7	0.9	543	5.8	6.7	0.9
506	7.3	7.8	0.5	545	7.4	7.9	0.5
524	6.8	7.3	0.5	553	7.4	8.1	0.7
529	8.1	8.0	–0.1	561	7.7	8.0	0.3
530	5.6	6.8	1.2	563	6.5	6.8	0.3

Preliminary Analysis

A preliminary check of the normality assumption for the *differences* (1996 – 1995) is shown below. Neither the plot nor the results of the normality test suggest that the assumption of normally distributed differences has been violated.



Formal Analysis

The test procedure is straightforward and is equivalent to a one-sample *t*-test performed on the differences. The degrees of freedom in this case are the number of *pairs* less one, not the total sample size less one.

t-Test of the Mean

```
Test of mu = 0.000 vs mu not = 0.000

Variable     N      Mean    StDev   SE Mean        t          P
diff        14     0.471    0.547     0.146     3.23     0.0066
```

Interpretation

At a 5% level, the p-value of 0.0066 is deemed to be significant and we therefore reject the hypothesis of equality of the two population means.

Worked Example 4: Fixed Effect ANOVA

Context

A biologist wanted to know if there was a difference in the abundance of the mayfly nymphs (*Ulmerophlebia* sp.) between the northern, southern, eastern, and western shores of Lake Windermere, Jervis Bay. The western shoreline bears the brunt of prevailing winds in the region, and the resulting wave action has reduced emergent macrophytic growth and accumulation of litter on both the western and the northern shores. Five replicate column collections of benthic invertebrates were collected from each of the four shores and the nymphs of the mayfly *Ulmerophlebia* sp. were counted. The data below have been transformed to a log scale to weaken the relationship between the variance and the mean.

West	North	East	South
0.48	1.11	2.00	1.94
0.78	1.15	1.36	1.52
1.11	1.57	1.86	1.60
0.60	0.95	2.33	1.81
1.30	1.48	1.83	2.26

Preliminary Analysis

Descriptive Statistics

```
Variable  direct          N     Mean   Median  Tr Mean    StDev  SE Mean
log(abund)W               5    0.854    0.778    0.854    0.346    0.155
          N               5    1.252    1.146    1.252    0.260    0.116
          E               5    1.876    1.857    1.876    0.349    0.156
          S               5    1.827    1.813    1.827    0.294    0.132

Variable  direct        Min      Max       Q1       Q3
log(abund)W           0.477    1.301    0.540    1.207
          N           0.954    1.568    1.034    1.523
          E           1.362    2.330    1.597    2.165
          S           1.519    2.260    1.560    2.100
```

Observe that the standard deviations are reasonably consistent over the four sites. The main difference in *log*-abundance seems to be depressed counts at the western and northern sites. The significance of these differences is formally examined via a one-way (fixed effects) ANOVA model.

Formal Analysis

```
Analysis of Variance for log(abund)
Source     DF        SS        MS        F        P
direct      3    3.5877    1.1959    12.09    0.000
Error      16    1.5830    0.0989
Total      19    5.1707
                                  Individual 95% CIs For Mean
                                  Based on Pooled StDev
Level       N      Mean     StDev  ---------+---------+---------+-------
W           5    0.8545    0.3459  (-----*-----)
N           5    1.2519    0.2596         (-----*-----)
E           5    1.8764    0.3495                      (-----*----)
S           5    1.8266    0.2942                     (-----*----)
                                  ---------+---------+---------+-------
Pooled StDev =   0.3145                 1.00      1.50      2.00
```

Interpretation

The computed *F*-ratio of 12.09 (with 3 and 16 degrees of freedom) is seen to be significant ($p < 0.0005$). Furthermore, the individual confidence intervals (based on the pooled estimate of common variance) suggest that abundances at the eastern and southern sites are similar as are the western and northern sites, but that these two groupings are different. The significance of this observation is examined via Tukey's multiple comparison technique.

Worked Example 4 (continued)

Tukey's pairwise comparisons

```
Family error rate = 0.0500
Individual error rate = 0.0113
Intervals for (column level mean) - (row level mean)
                W           N           E
       N     -0.9672
              0.1722
       E     -1.5916     -1.1942
             -0.4522     -0.0548
       S     -1.5419     -1.1444     -0.5199
             -0.4024     -0.0050      0.6195
```

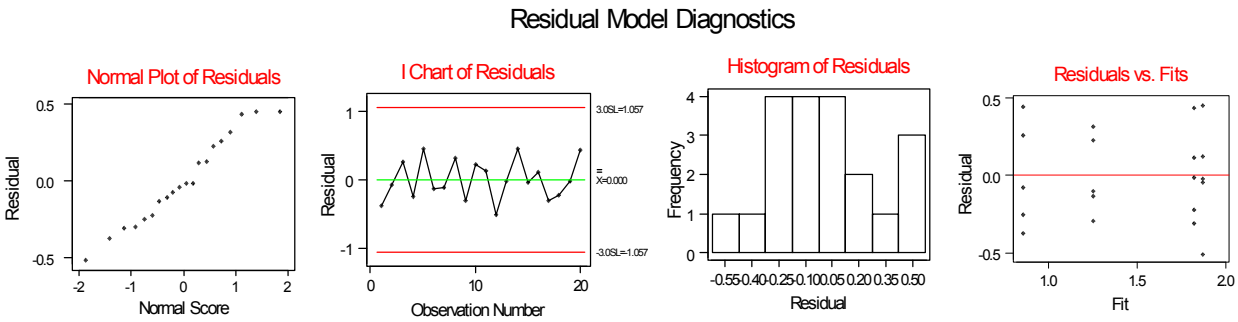
Interpretation

The analysis above shows a matrix of confidence limits for the respective comparison between row and column means. For example, the top left cell shows that the confidence interval for the *difference* between the western and northern site means extends from –0.9672 to 0.1722. Given that equality of the two means corresponds to a *zero* difference we would *not* reject a null hypothesis of no difference between this pair of sites (since zero is spanned by the confidence interval). Similar interpretations may be made for the other cell entries. Overall, our conclusion is as we suspected — there is no difference between North–West or East–South sites, but these groupings are significantly different from each other. The so-called 'Family error rate' is effectively the Type I error for the set of six comparisons — that is, the probability of incorrectly rejecting at least one true null hypothesis of no difference is 0.05. To achieve this overall error rate of 0.05, each individual comparison has a Type I error of 0.0113.

Residual Analysis

It is always instructive to look at the *residuals* (i.e. the portion of the original measurement that the statistical model has not been able to account for) after the ANOVA model has been fitted. A number of diagnostic tools are available to help decide if there have been any violations of the attendant ANOVA assumptions.

The plots below show one such set of diagnostic tools:



The normal probability plot shows some departure from linearity (particularly in the tails of the distribution) suggesting that the normal assumption may have been violated (further testing as in Worked Example 1 would be necessary to establish the significance of these departures). The I-chart shows the residuals plotted in sequential order. Evidence of trends or non-random behaviour is indicative of some underlying process (e.g. seasonal effects) that has not been captured by the model and may invalidate the ANOVA output. Such effects do not appear to be evident from the plot. The histogram provides another visual check on the distribution of residuals. The non-normality is apparent, although the small sample sizes should be kept in mind when assessing the histogram. Given that the normality assumption is a robust assumption, the shape of the histogram of residuals should not overly concern us. Finally, the plot of residuals against the fitted values (i.e. the predicted mean abundance in each of the four directions) should also show no systematic trends, drifts, periodicities or other non-random behaviour. We should also look for evidence of gross discrepancies in the degree of spread between the different factor levels as evidence of non-constant variance. In the example above, no particular concerns are raised.

Worked Example 5: Single Factor ANOVA Planned Comparison

Context

Phosphorus is an important nutrient in aquatic ecosystems; concentrations can be changed dramatically through non-natural discharges into streams and lakes. The following data are for concentrations of phosphorus (μg/L) in samples of water taken at various distances up- and downstream of a waste water outlet. The data below are replicates taken from their respective locations at the one time:

Distance downstream (km)				
–0.5	0.0	1.0	2.0	3.0
4.86	6.16	6.82	5.86	5.31
4.86	5.83	6.67	5.73	4.98
5.19	6.93	6.34	5.62	4.98
4.31	6.16	6.08	4.83	5.46
4.99	6.93	5.73	5.49	4.66

The water scientist wanted to know whether the mean phosphorus concentrations of sites downstream from the effluent outlet differed from the upstream site ('control'). The scientist also needed to assess if the impact, if any, could be considered local or if it persisted well downstream.

Analysis

Analysis of Variance for P

```
Source     DF        SS        MS        F        P
Dist        4    10.121     2.530    15.59    0.000
Error      20     3.246     0.162
Total      24    13.367
```

```
                                Individual 95% CIs For Mean
                                Based on Pooled StDev
Level       N      Mean     StDev  -------+---------+---------+---------
-0.5        5    4.8420    0.3266   (----*-----)
 0.0        5    6.4020    0.5005                        (----*-----)
 1.0        5    6.3280    0.4411                       (----*-----)
 2.0        5    5.5060    0.4018            (-----*----)
 3.0        5    5.0780    0.3137      (-----*----)
                                -------+---------+---------+---------
Pooled StDev =   0.4029              4.90      5.60      6.30
```

Interpretation

The ANOVA output suggests a highly significant 'distance' effect — in other words, phosphorus concentrations are related to distance from outlet. Examination of the individual confidence intervals suggests upstream concentrations are reached beyond 2 km downstream from the outlet. Marked differences in concentrations are observed at the outlet and 1 km downstream.

A formal test of the significance of these observations can be made through the use of an appropriate multiple comparison procedure. Tukey's method was used in Worked Example 4, although the situation here is slightly different because one of the locations is a control with which all other sites are to be compared. The appropriate procedure in this case is Dunnett's test. Like Tukey's test, Dunnett's test keeps the overall experiment-wise error rate fixed at the nominal 0.05 level by conducting each of the individual comparisons at an appropriately smaller level of significance.

Dunnett's intervals for treatment mean minus control mean

```
Family error rate = 0.0500
Individual error rate = 0.0153
Control = level (-0.5) of Dist
Level        Lower    Centre     Upper ------+---------+---------+---------+-
 0.0        0.8848    1.5600    2.2352                  (-------*--------)
 1.0        0.8108    1.4860    2.1612                 (--------*-------)
 2.0       -0.0112    0.6640    1.3392       (-------*--------)
 3.0       -0.4392    0.2360    0.9112  (-------*-------)
                                      ------+---------+---------+---------+-
                                         0.00      0.80      1.60      2.40
```

Our intuition is supported by the results of Dunnett's test. We see that the control site is significantly different from the 0.0 and 1.0 downstream sites (since zero is not encompassed by the relevant interval) but is not significantly different from the 2.0 or 3.0 downstream sites.

Worked Example 6: Nested Analysis of Variance

Context

A water scientist was interested to know if there was a difference in fluoride concentration between three locations. Three independent water samples were taken from each location. Two independent determinations of fluoride content (mg/L) were made on each sample. The data were as follows (data from Zar 1984):

Locations	1			2			3		
Samples	1	2	3	1	2	3	1	2	3
	1.1	1.3	1.2	1.3	1.3	1.4	1.8	2.1	2.2
	1.2	1.1	1.0	1.4	1.5	1.2	2.0	2.0	1.9

This is in the design of a nested ANOVA with samples (random factor) nested within location (fixed factor). The inclusion of a random-effects nested factor (sample) enables us to account for some of the within-group variability, and hence improves the sensitivity of the test with respect to location effects.

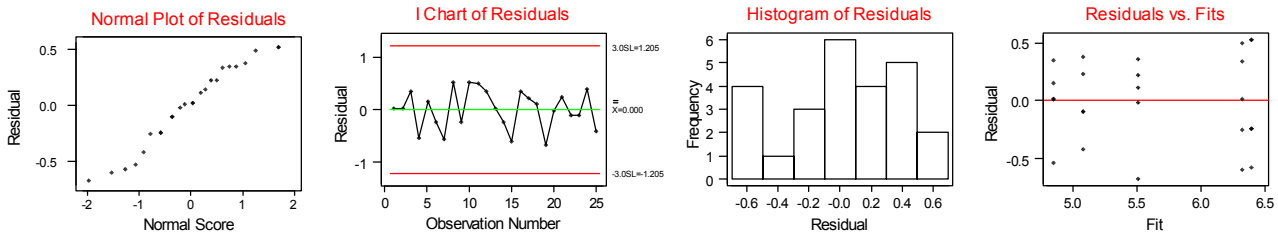
Analysis

This nested ANOVA model is conveniently specified using the GLM (General Linear Model) option available in most statistical software packages. The following output is obtained.

General Linear Model

```
Factor              Type  Levels   Values
Location           fixed       3   1 2 3
Sample(Location) random       9   1 2 3 1 2 3 1 2 3

Analysis of Variance for Concen, using Adjusted SS for Tests

Source              DF     Seq SS     Adj SS     Adj MS       F      P
Location             2    2.37000    2.37000    1.18500  142.20  0.000
Sample(Location)     6    0.05000    0.05000    0.00833    0.47  0.816
Error                9    0.16000    0.16000    0.01778
Total               17    2.58000

Variance Components, using Adjusted SS

Source            Estimated Value
Sample(Location)          -0.00472
Error                      0.01778
```

An analysis of the residuals after fitting the nested ANOVA model is provided below.



Interpretation

The ANOVA output suggests a highly significant 'location' effect, although samples within location do not differ significantly.

With respect to the analysis of residual diagnostics: there is some departure from normality (although not serious) in the lower tail of the distribution (refer normal probability plot and histogram). The plot of residuals versus fitted values shows some evidence of non-constant variance. Formal tests of the homogeneity of variance assumption are available (see Worked Example 9, for instance).

Worked Example 7: Equality of Variances

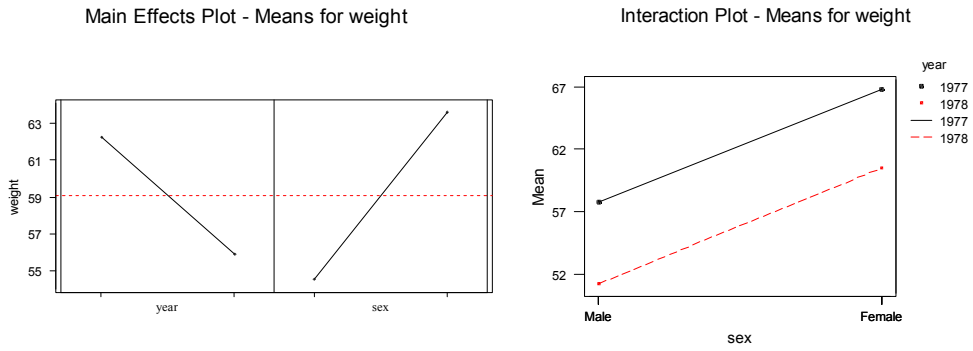
Context

A researcher was interested in determining if there was a difference in body weight of *Antechinus stuartii* between the years 1977 and 1978. The researcher knew that females were bigger than males so the sex of each individual was also recorded. The data were as follows:

	1977	1978		1977	1978
	56	52		61	54
Male	56	52	Female	63	64
	62	52		72	64
	57	49		71	60

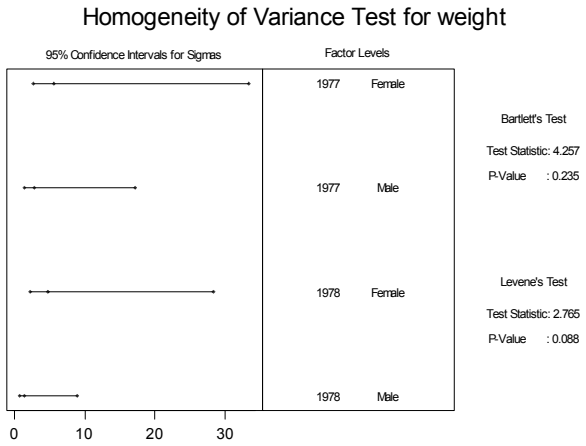
Analysis

It is always useful to inspect main effects and interaction plots in multi-factor designs to gain a better appreciation of the results of any subsequent ANOVA / GLM analysis.



The plots above suggest the presence of both a 'year' and a 'sex' main effect. The interaction plot shows two almost parallel lines that are indicative of an absence of any interaction between the main effects. The significance of these observations can be formally tested via a two-factor ANOVA.

Perhaps more important than the assumption of normality, is the ANOVA assumption of equality of group variances. It is good practice to examine the group statistics and make some preliminary assessment of the applicability of the homogeneity of variances assumption. More formal statistical tests are available, two of which (Bartlett's and Levene's test) are reproduced below. It should be noted that one drawback of Bartlett's test is its sensitivity to departures from normality. A more robust test is Hartley's F_{max} test (not conducted here).



Interpretation

In this case, we accept the hypothesis of equal group variances at the 5% level, although the wide discrepancy in *p*-values between the two test procedures is noted. Good statistical practice would dictate that the most appropriate test is identified and then applied to the data. To apply a battery of tests and choose the most appealing result is of course unscientific and negates the application of any statistical method.

Worked Example 8: Two-factor ANOVA

Context

As for Worked Example 7.

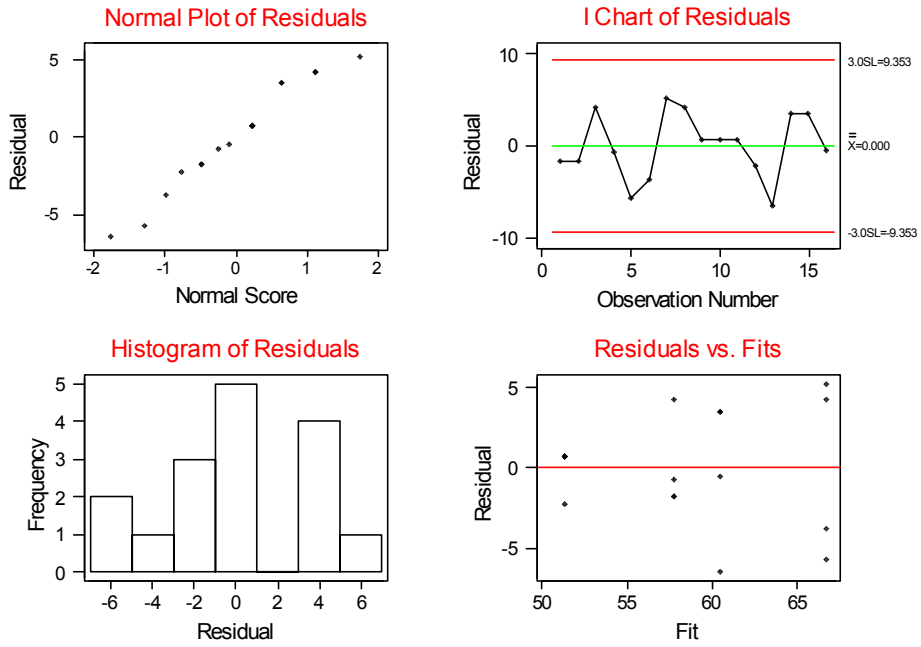
Analysis

A fixed two-factor analysis of variance is appropriate as it enables the comparison of weight between years after adjusting for differences between sexes. Conversely sex can also be compared after adjusting for differences between years.

```
Analysis of Variance for weight

Source      DF     Seq SS     Adj SS     Adj MS       F      P
year         1     162.56     162.56     162.56   10.20  0.008
sex          1     333.06     333.06     333.06   20.90  0.000
year*sex     1       0.06       0.06       0.06    0.00  0.951
Error       12     191.25     191.25      15.94
Total       15     686.94
```

Interpretation

This analysis confirms our observations. Both main effects are significant at a 5% level, while the interaction is non-significant. The residual diagnostics appear below.



The residual diagnostics give no cause for concern or remedial action.

Worked Example 9: Analysis of Covariance

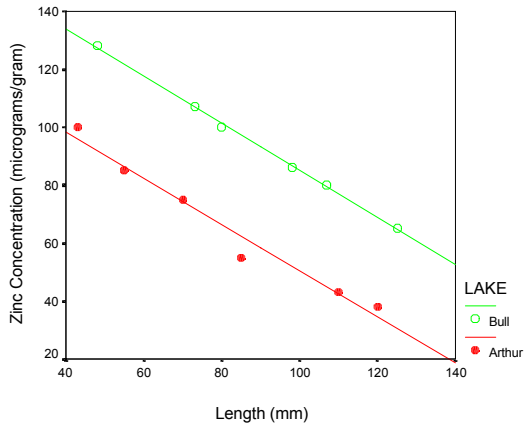
Context

Fisheries were interested to know whether fish in a polluted environment were likely to accumulate higher concentrations of heavy metals than fish in a more pristine habitat. To investigate this, six flathead were caught in both Lake Bull (polluted urban lake) and Lake Arthur (relatively pristine) and analysed for zinc (µg/g). As heavy metal concentration in organisms is often related to body size, the lengths of the fish were also recorded. The data were as follows:

Site	Zinc	Length	Site	Zinc	Length
Lake Arthur	43	110	Lake Bull	80	107
Lake Arthur	75	70	Lake Bull	86	98
Lake Arthur	38	120	Lake Bull	65	125
Lake Arthur	55	85	Lake Bull	100	80
Lake Arthur	100	43	Lake Bull	128	48
Lake Arthur	85	55	Lake Bull	107	73

As there was large variability in the size of fish sampled, it was decided that comparisons of zinc concentrations between lakes should be made by analysis of covariance (ANCOVA). The ANCOVA tests for differences in zinc concentration in fish between lakes after correcting the zinc measurements for the effect of fish body size.

Analysis



- We first determine that the relationship between the dependent variable (Zinc) and the covariate (fish length) is linear for both lakes.
- From the diagram it is clear that there is a strong negative linear relationship between zinc concentration and the length of fish for both Lake Bull and Lake Arthur. Therefore, zinc concentration is dependent on fish size in both lakes.

The zinc data are analysed using an ANCOVA model with length being a covariate. We note that two model formulations are possible, depending on the inference to be made about the covariate.

In the first formulation we use a nested model to force the explicit estimation of separate regression slopes (i.e. one regression for each of the two lakes).

Nested Analysis of Variance for Zinc

```
Source          DF     Seq SS     Adj SS     Adj MS       F      P
Lake             1     2408.3      337.3      337.3   30.15  0.000
Length(Lake)     2     5403.9     5403.9     2701.9  241.57  0.000
Error            8       89.5       89.5       11.2
Total           11     7901.7

Term                 Coef     StDev        T      P
Constant          148.298     3.278    45.24  0.000
Length (Lake)
       Arthur   -0.79870   0.04922   -16.23  0.000
       Bull     -0.81318   0.05485   -14.83  0.000
```

Worked Example 9 (continued)

```
Unusual Observations for Zinc

Obs      Zinc       Fit   StDev Fit  Residual   St Resid
  4    55.000    62.406       1.383    -7.406     -2.43R

R denotes an observation with a large standardised residual
```

From this analysis we conclude that (i) zinc concentrations are significantly different between the two lakes, and (ii) there is a significant relationship (the implied null hypothesis is that of zero slope for the regression) between zinc levels and length in each of the two lakes. The computer output has also flagged a potential outlier or in some other way aberrant observation (Observation #4). The high standardised residual for this observation is a consequence of the extremely high zinc reading.

The next model formulation is a fully crossed design in which the overall regression effect is estimated and its significance tested.

Fully-crossed Analysis of Variance for Zinc

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Lake	1	2408.3	337.3	337.3	30.15	0.000
Length	1	5403.4	5350.8	5350.8	478.39	0.000
Lake*Length	1	0.4	0.4	0.4	0.04	0.849
Error	8	89.5	89.5	11.2		
Total	11	7901.7				

Term		Coef	StDev	T	P
Constant		148.298	3.278	45.24	0.000
Length		-0.80594	0.03685	-21.87	0.000
Length*	Lake				
	Arthur	0.00724	0.03685	0.20	0.849

Interpretation

From this analysis we conclude that the overall coefficients of the regression slope and intercept are both highly significant (i.e. non-zero). Furthermore, the interaction term is non-significant suggesting that separate slopes are not warranted (implying essentially parallel lines).

Worked Example 10: Generalised Linear Models

Context

We note that the data analysed in this example take the form of counts, which are necessarily discrete. By back-transforming, the original set of abundances is obtained. These are shown below:

Variable	direct	N	
Count	W	5	3, 6, 13, 4, 20
	N	5	13, 14, 37, 9, 30
	E	5	100, 23, 72, 214, 68
	S	5	87, 133, 40, 65, 182

It is evident that there are not only substantial differences *between* directions, but also of the abundances *within* each direction. Generally, normal-based models are satisfactory for the analysis of count data provided the average counts are reasonably large (typically > 30). While this is mostly true for the data above, some quite low counts are observed in the north and westerly directions.

Analysis

A more flexible, and conceptually more appropriate approach to the analysis of this type of (discrete) data is afforded by a *generalised linear model* having a Poisson error term and *log-link*. The details of this approach are not covered here, although the analysis is similar to a conventional ANOVA with the exception that we look at changes to the *deviance* statistic as a series of nested models is fitted.

Interpretation

For the original data, we find the null model (corresponding to a single overall mean) has a deviance of 1019.3 with 19 degrees of freedom. By adding the 'direction' factor to the model, this deviance is reduced to 411.71 (16 degrees of freedom). The change in deviance 607.59 is highly significant when compared to a chi-squared test statistic having 3 (19–16) degrees of freedom.

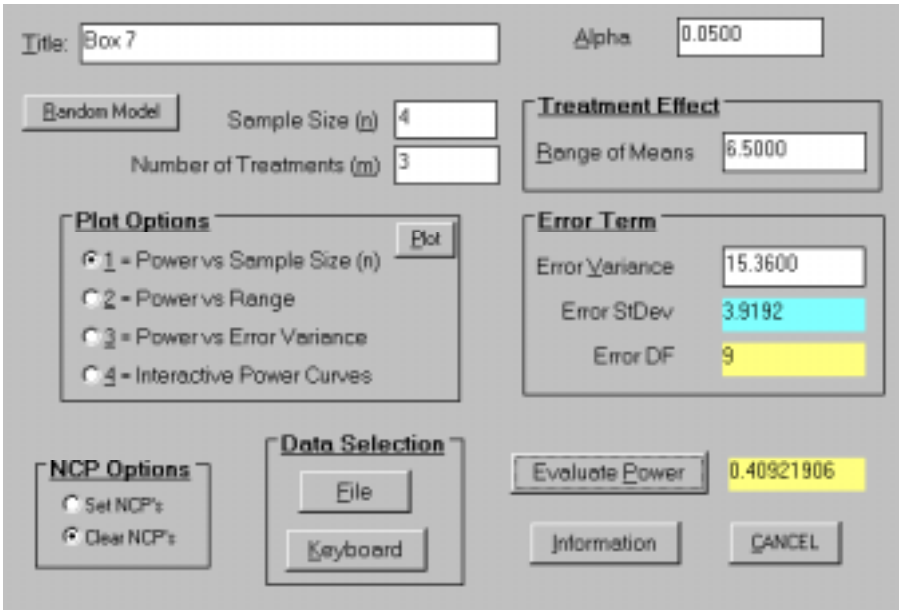
Worked Example 11: Power and Sample Size Determinations

Context

A pilot study was carried out to determine the mean cadmium concentrations in the cockle *Anadara trapezia* from each of three sites. Estimates of means for each site from four samples were: A: 12.75; B: 17.75; C: 19.25, and σ^2 was estimated to be 15.36. What will be the power of the ANOVA in detecting a difference between sites if we test at the 0.05 level of significance?

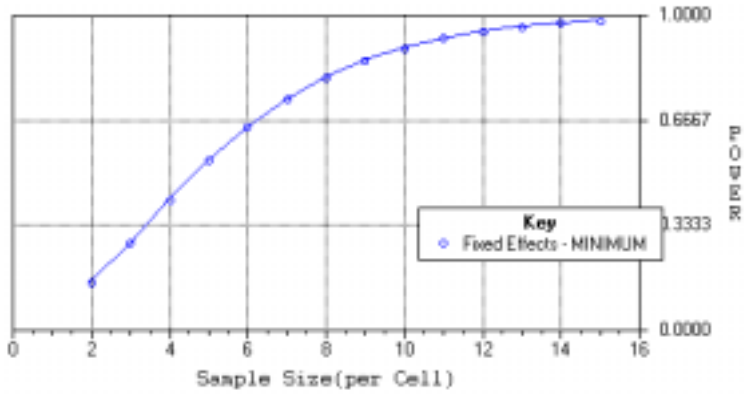
Analysis

Output from the *PowerPlant*® software (see section A5.1.10 for download details) is shown below.



Interpretation

We see immediately from the 'Evaluate Power' box, that the estimated power for this design is 0.409. Sample size determinations are readily assessed from the power curve.



We see that a sample size of about 13 per treatment group should give the required 0.8 power.

Worked Example 12: Regression

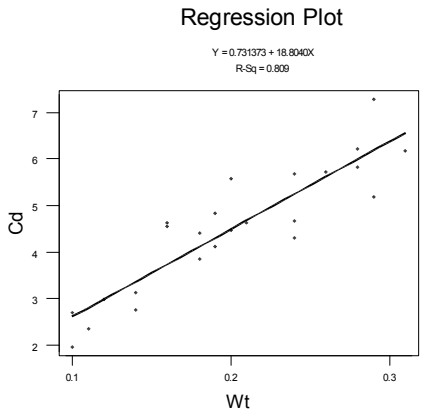
Context

A marine scientist wanted to determine if the cockle *Anadara trapezia* was an appropriate bioindicator for heavy metal pollution. Other species of bivalves have previously been found to accumulate heavy metals in their body tissues. Therefore, 24 cockles were dried, weighed, and the cadmium determined.

Cd	Wt	Cd	Wt	Cd	Wt	Cd	Wt
2.71	0.10	2.35	0.11	1.97	0.10	4.67	0.24
2.76	0.14	4.46	0.20	4.54	0.16	3.14	0.14
4.12	0.19	4.64	0.16	4.41	0.18	5.57	0.20
4.83	0.19	3.85	0.18	5.68	0.24	5.72	0.26
6.16	0.31	6.21	0.28	4.30	0.24	4.64	0.21
7.27	0.29	3.00	0.12	5.82	0.28	5.18	0.29

Issue to be addressed: Is there a relationship between the size of the cockle and the total amount of cadmium in the tissues? A positive relationship would suggest that cockles do in fact accumulate cadmium in their tissues over time.

Analysis

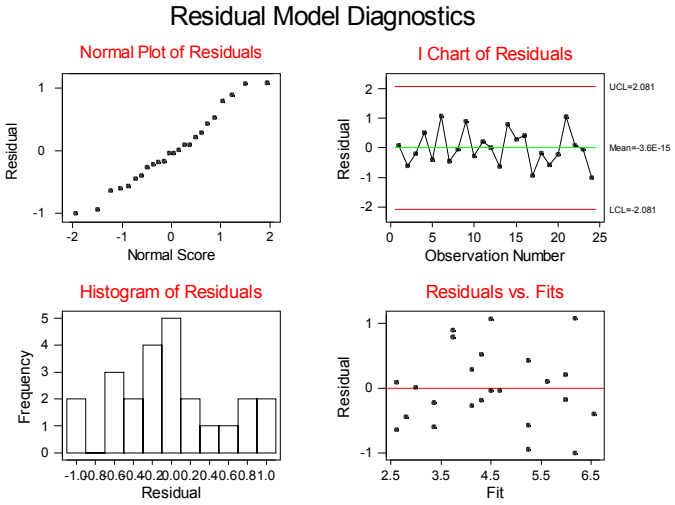



```
The regression equation is
y = 0.731 + 18.8 x

Predictor       Coef     StDev         T       P
Constant      0.7314    0.4091      1.79   0.088
x             18.804     1.947      9.66   0.000

S = 0.6018     R-Sq = 80.9%     R-Sq(adj) = 80.0%
```

Interpretation

Both the plot and the analysis indicate a strong, positive linear relationship between Cd and weight. Inspection of the *p*-values associated with terms in the regression model suggests that the intercept (0.7314) is not significantly different from zero (hence a regression model through the origin might be contemplated) while the regression slope (18.804) is significantly different from zero.

An important follow-up analysis relates to the regression diagnostics. These appear below. These diagnostic plots raise no particular concerns.



Appendix 6

Typical Field Record Sheet & Laboratory Request Form

A6.1. Field Record Sheet

Officer/s……..…….....……. **Date**……

Sampling run number:……..... **Site code**:…….....…….

Site name…….....…

Time: start..... **finish**.....

Field measurements:

Parameter	Result
Depth (m)	
Secchi depth (m)	
Altitude (m)	
Temperature (^{o}C)	
Turbidity (NTU)	
Dissolved oxygen (mg/L) (% saturation)	
Electrical conductivity (mS/cm)	
pH	
Salinity ( )	
Eh (mV)	
Others	

Field observations:

Station no.…….

Description.....

.....

Observation	Details
Weather: e.g. wind, wind direction, cloud cover	
Colour and appearance of water	
Water surface condition	
Water flow, level, tide:	
Presence of nuisance organisms (e.g. macrophytes, phytoplankton scums, algal mats)?	
Presence of oily films on surface or on shoreline?	
Presence of floating debris or grease?	
Presence of odour or frothing?	
Other observations	

Signature.....

(when sample collected and entries completed)

Water quantity measurement data:

Location description.....

Description of gauge.....

Stage height.....

Time.....

Sample details:

Analyte	Container material	Volume collected	Preservation	Quality control
Major ions				
Metals				
Organic compounds				
Pesticides and herbicides				
Mercury				
Phenols				
Nutrients				
BOD and COD				
Others				

Quality Control Remarks:.....

.....

.....

.....

.....

.....

.....

A6.2. Typical Laboratory Request Form

(partially filled-in as an example)

Sample program / site description.....

Sampling officer.....**Title**.....**Section**.....**Branch**.....

Sampling date.....**File Reference**.....**Payment Authority No.**.....

Sample no.	Sample time	Sample location	Sample description	Parameters to be analysed	Sample size (mL or g)	Reference criteria
MB1	09.00	Monitor bore (north-east)	Groundwater	pH, EC, TP, TN	1000 mL	SW, IP
SP2	09.45	Creek upstream of site	Creek water	Colour, NFR, EC, TP, TN	1000 mL	A, AE

Legend for reference criteria: AE = aquatic ecosystems, **DW** = drinking water, **IP** = irrigation of plants, **SW =** stock water supplies, **IW** = industrial water supplies, **A** = aesthetic values; **R =** Recreational waters (reference: ANZECC *Australian & NZ Water Quality Guidelines for Fresh and Marine Waters* (1992))

Site conditions during sampling.....
(e.g. Cool 12^{o}C, raining, creek flowing at estimated 10–15 L/sec)

Sample preservation details.....(e.g. On ice)

Analytical laboratory.....**Accepted by**.....**Title**.....

Date.....**Time**.....

Comments:

Analysis request notes (tick requirement)**:**	**Return analysis results to:**.....
Routine..... (.....) — mail results when ready Urgent..... (.....) — fax results to ....., as soon as possible Legal action.....(.....) — ensure chain of custody and data validation	Entity..... Address..... ....

References

Abdullah, A.R., Kumar, A. and Chapman, J.C. (1994): Inhibition of acetylcholinesterase in the Australian freshwater shrimp (*Paratya australiensis*) by profenofos. *Environ. Toxicol. Chem. 13:* 1861–1866.

Adams, M.S. (2000): Development of a whole sediment toxicity test using a benthic marine microalga. Honours Thesis, University of Technology, Sydney.

Ahlers, W.W., Reid, M.R., Kim, J.P. and Hunter, K.A. (1990): Contamination-free sample collection and handling protocols for trace elements in natural fresh waters. *Aust. J. Mar. Freshwater Res. 41:* 713–720.

ANZECC & ARMCANZ (1994): *Policies and Principles — A Reference Document.* National Water Quality Management Strategy Paper no. 2. Australian and New Zealand Environment and Conservation Council / Agriculture and Resource Management Council of Australia and New Zealand.

ANZECC & ARMCANZ (1998): *Implementation Guidelines.* National Water Quality Management Strategy Paper no. 3. Australian and New Zealand Environment and Conservation Council / Agriculture and Resource Management Council of Australia and New Zealand.

ANZECC & ARMCANZ (2000): *Australian and New Zealand Guidelines for Fresh and Marine Water Quality*. 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.

ANZECC (1998): *Core Environmental Indicators for Reporting on the State of the Environment — Discussion Paper for Comment*. State of the Environment Reporting Task Force. Australian and New Zealand Environment and Conservation Council Secretariat, Canberra.

At URL http://www.environment.gov.au/soe/envindicators/coreindicators.html

APHA (1998): *Standard Methods for the Examination of Water and Wastewater*. 20th Edition, American Public Health Association, Water Environment Federation, American Water Works Association, Washington, USA.

Apte, S.C., Batley, G.E., Szymczak, R., Rendell, P.S., Lee, R. and Waite, T.D. (1998): Baseline trace metal concentrations in New South Wales coastal waters. *Mar. Freshwater Res. 49:* 203–214.

Arthington, A.H., Kennard, M.J. and Mackay, S.J. (1998): River bioassessment using fish community structure. Report GRU19 to the Land and Water Resources Research and Development Corporation.

AS (1987a): Australian Standard AS2031.1-1987. *Selection of Containers and Preservation of Water Samples for Chemical and Microbiological Analysis. Part 1. Chemical Analysis*. Standards Australia, Homebush, NSW.

AS (1987b): Australian Standard AS2031.1-1987. *Selection of Containers and Preservation of Water Samples for Chemical and Microbiological Analysis. Part 2. Microbiological Analysis*. Standards Australia, Homebush, NSW.

AS (1990): Australian Standard AS3550.4 *Methods for the Analysis of Waters. Determination of Solids — Gravimetric Method*. Standards Australia, Homebush, NSW.

AS/NZS (1995): Australian/New Zealand Standard AS/NZS 2243.3:1995. *Safety in Laboratories — Part 3: Microbiology*. Standards Australia, Homebush, NSW.

AS/NZS (1997a): Australian/New Zealand Standard AS/NZS 2243.1:1997. *Safety in Laboratories — Part 1: General*. Standards Australia, Homebush, NSW.

AS/NZS (1997b): Australian/New Zealand Standard AS/NZS 2243.2:1997. *Safety in Laboratories — Part 2: Chemical Aspects.* Standards Australia, Homebush, NSW.

AS/NZS (1998a): Australian/New Zealand Standard AS/NZS 5667.1:1998. *Water Quality — Sampling — Guidance on the Design of Sampling Programs, Sampling Techniques and the Preservation and Handling of Samples*. Standards Australia, Homebush, NSW.

AS/NZS (1998b): Australian/New Zealand Standard AS/NZS 5667.9:1998. *Water Quality — Sampling — Guidance on Sampling From Marine Waters*. Standards Australia, Homebush, NSW.

AS/NZS (1998c): Australian/New Zealand Standard AS/NZS 5667.6:1998. *Water Quality — Sampling — Guidance on Sampling of Rivers and Streams*. Standards Australia, Homebush, NSW.

AS/NZS (1998d): Australian/New Zealand Standard AS/NZS 5667.4:1998. *Water Quality — Sampling — Guidance on Sampling from Lakes, Natural and Man-Made*. Standards Australia, Homebush, NSW.

AS/NZS (1998e): Australian/New Zealand Standard AS/NZS 5667.11:1998. *Water Quality — Sampling — Guidance on Sampling of Groundwaters*. Standards Australia, Homebush, NSW.

AS/NZS (1999): Australian/New Zealand Standard AS/NZS 5667.12:1998. *Water Quality — Sampling — Guidance on Sampling of Bottom Sediments*. Standards Australia, Homebush, NSW.

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.

Barbour, M.T., Plafkin, J.L., Bradley, B.P., Graves, C.G. and Wisseman, R.W. (1992): Evaluation of EPA's rapid bioassessment benthic metrics: metric redundancy and variability among reference stream sites. *Environmental Toxicology and Chemistry 11:* 437–449.

Barbour, M.T., Stribling, J.B. and Karr, J.R. (1995): Multimetric approach for establishing biocriteria and measuring biological condition. *Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making* (W.S. Davis and T.P. Simon, eds), pp. 63–77. Lewis, Boca Raton, FL, USA.

Barbour, M.T., Gerritsen, J., Griffith, G.E., Frydenborg, R., McCarron, E., White, J.S. and Bastian, M.L. (1996): A framework for biological criteria for Florida streams using benthic macroinvertebrates. *Journal of the North American Benthological Society 15:* 185–211.

Barceló, C., Pawlowsky, V. and Grunsky, E. (1996): Some aspects of transformations of compositional data and the identification of outliers. *Mathematical Geology 28:* 501–518.

Bardwell, D.A. (1991): Problem-framing: a perspective on environmental problem solving. *Environmental Management 15:* 603–612.

Batley, G.E. (1989): Collection, preparation and storage of samples for speciation analysis. In *Trace Element Speciation: Analytical Methods and Problems* (G.E. Batley, ed.), pp. 1–24. CRC Press Inc. Florida.

Belbin, L. (1993): *PATN — Pattern Analysis Package*. CSIRO Division of Wildlife and Rangelands Research, Canberra.

Belbin, L. and McDonald, C. (1993): Comparing three classification strategies for use in ecology. *Journal of Vegetation Science 4:* 341–348.

Bennett, C.A. and Franklin, N.L. (1988): *Statistical Analysis in Chemistry and the Chemical Industry.* Qualpro Inc.

Beyers, D.W. (1998): Causal inference in environmental impact studies. *Journal of the North American Benthological Society 17:* 367–373.

Blomqvist, S. (1991): Quantitative sampling of soft-bottom sediments: problems and solutions. *Mar. Ecol. Prog. Ser. 72:* 295–304.

Box, G.E.P. and Cox, D.R. (1964): An analysis of transformations (with discussion). *Journal of the Royal Statistical Society Series B 26:* 211–246.

Box, G.E.P. and Cox, D.R. (1982): An analysis of transformations revisited, rebutted. *Journal of the American Statistical Association 77:* 209–210.

Brodie, J. (1996): *River Flood Plumes in the Great Barrier Reef.* Department of Earth Sciences JCU, & CRC Research.

Brodie, J. and Furnas, M. (1996): Cyclone, river flood plumes and natural water quality extremes in the central Great Barrier Reef. In *Downstream Effects of Landuse*. (H.G. Hunter, A.G. Eyles, and G.E. Rayment, eds), pp. 367–374. Department of Natural Resources, Queensland, Australia.

Bunn, S.E., Davies, P.M. and Kellaway, D.M. (1997): Contributions of sugar cane and invasive pasture grass to the aquatic food web of a tropical lowland stream. *Mar. Freshwater Res. 48:* 173–179.

Bunn, S.E., Davies, P.M. and Mosisch, T.D. (1999): Ecosystem measures of river health and their response to riparian and catchment degradation. *Freshwater Biology 41:* 333–345.

Burridge, T., Portelli, T. and Ashton, P. (1996): Effect of sewage on germination of three marine brown algal macrophytes. *Mar. Freshwater Res. 47:* 1009–1014.

Burridge, T.R., Lavery, T. and Lam, P. (1995): Acute toxicity tests using *Phyllospora comosa* (Labillardiere) C. Agardh, (Phaeophyta: Fucales) and *Allorchestes compressa* Dana (Crustacea: Amphipoda). *Bulletin for Environmental Contamination and Toxicology 55:* 621–628.

Carpenter, S.R., Frost, T.M., Heisey, D. and Kratz, T.K. (1989): Randomized intervention analysis and the interpretation of whole-ecosystem experiments. *Ecology 70:* 1142–1152.

Chambers, J.M. and Hastie, T.J. (Eds) (1992): *Statistical Models in S*. Wadsworth and Brooks, CA, USA.

Chapman, P.M. (1990): The Sediment Quality Triad approach to determining pollution-induced degradation. *Sci. Tot. Environ. 97-8:* 815–825.

Chatfield, C. (1980): *Introduction to Multivariate Analysis*. Chapman and Hall, London.

Chessman, B.C. (1995): Rapid river assessment using macroinvertebrates: a procedure based on habitat-specific family level identification and a biotic index. *Aust. J. Ecol. 20:* 122–129.

Chessman, B.C. and McEvoy, P.K. (1998): Towards diagnostic biotic indices for river macroinvertebrates. *Hydrobiol. 364:* 169–182.

Chessman, B.C., Growns, J.E. and Kotlash, A.R. (1997): Objective derivation of macroinvertebrate family sensitivity grade numbers for the SIGNAL biotic index: application to the Hunter River system, New South Wales. *Mar. Freshwater Res. 48:* 159–172.

Chessman, B., Growns, J., Currey, J. and Plunkett-Cole, N. (1999): Prediction of diatom communities at the genus level for the rapid assessment of rivers. *Freshwater Biol. 41:* 317–331.

Clarke, K.R. and Warwick, R.M. (1994): *Change in Marine Communities: An Approach To Statistical Analysis and Interpretation*. Plymouth Marine Laboratories.

Cleveland, W.S. (1979): Robust locally-weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc. 74:* 829–836.

Cochran, W.G. (1977): *Sampling Techniques*, 3rd edition. John Wiley, New York, USA.

Cohen, J. (1988): *Statistical Power Analysis for the Behavioural Sciences*, 2nd edn. Lawrence Earlbaum Associates, Hillsdale, NJ.

Collier, K.J., Wilcock, R.J. and Meredith, A.S. (1998): Influence of substrate type and physico-chemical conditions on macroinvertebrate faunas and biotic indices of some lowland Waikato, New Zealand, streams. *New Zealand Journal of Marine and Freshwater Research 32:* 1–19.

Connell, D.W. (1981): *Water Pollution: Causes and Effects in Australia and New Zealand*, 2nd edn. University of Queensland Press, St Lucia.

Coughanowr, C. (1995): *The Derwent Estuary Nutrient Program Technical Report*. September 1995. Department of Environment and Land Management, Hobart.

Coughanowr, C. (1997): *State of the Derwent Estuary: A Review of Environmental Quality Data to 1997*. Supervising Scientist Report 129. Supervising Scientist, Canberra.

Coysh, J.L., Nichols, S.J., Simpson, J.C., Norris, R.H., Barmuta, L.A., Chessman, B.C. and Blackman, P. (2000): http://ausrivas.canberra.edu.au/, Web page for the AUStralian RIVer Assessment System (AUSRIVAS).

Cressie, N.A.C. (1993): *Statistics for Spatial Data*, Revised edn. John Wiley and Sons, New York, NY, USA.

Crompton, T.R. (1989): *Analysis of Seawater*. Butterworths, London.

Cryer, J. (1986): *Time Series Analysis*. Duxbury Press, Boston.

Cummins, K.W. and Klug, M.J. (1979): Feeding ecology of stream invertebrates. *Annual Review of Ecology and Systematics 10:* 147–172.

Davies, P.E., Fulton, W. and Kalish, S. (1988): *The Environmental Effects of Effluent from the ANM Newsprint Mills at Boyer, Tasmania*. Inland Fisheries Commission.

Davies, P.E., Fulton, W. and Kalish, S. (1989): *Water Quality in the Upper Derwent Estuary, Tasmania*. Inland Fisheries Commission Occasional Report No. 89-03.

Davies, P.M. (1997): *Assessment of River Health By Measurement of Community Metabolism*. Report UWA14 to the Land and Water Resources Research and Development Corporation.

Deeley, D.M. and Paling, E.I. (1999): *Assessing the Ecological Health of Estuaries*. LWRRDC Occasional paper 17/99. Urban sub program Report No. 10.

DeShon, J.E. (1995): Development and application of the invertebrate community index (ICI). *Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making* (W.S. Davis and T.P Simon, eds), pp. 217–244. Lewis, Boca Raton, FL, USA.

DEST (1995): *Water Quality Monitoring in Australia*. Department of Environment, Sport and Territories, Commonwealth of Australia.

Devries, D.R. and Stein, R.A. (1991): Comparison of three zooplankton samplers: a taxon-specific assessment. *J. Plankton Res. 13:* 53–59.

Dixon, W. and Chiswell, B. (1996): Review of aquatic monitoring program design. *Water Research 30:* 1935–1948.

Dobson, A.J. (1990): *An Introduction to Generalised Linear Models*. Chapman and Hall.

Donohue, R. (2000 unpublished): Ecosystem excursion from trigger values: the binomial approach. Available from author; see section 6.4.3.6 for details.

Doyle, C. (1998): Toxicity characterisation of contaminated sediment pore water using the sea urchin *Heliocidaris tuberculata*. Honours thesis, University of Technology, Sydney.

EA/NLWRA (2000): *Water Monitoring in Australia.* Report and database. Environment Australia, and National Land and Water Resources Audit, Canberra.

Eberhardt, L.L. (1976): Quantitative ecology and impact assessment. *J. Environ. Management 4:* 213–217.

Eberhardt, L.L. (1978): Appraising variability in population studies. *J. Wildlife Management 42:* 207–238.

Eberhardt, L.L. and Thomas, J.M. (1991): Designing environmental field studies. *Ecological Monographs 61:* 53–73.

Elliott, J.M. (1977): *Some Methods for the Statistical Analysis of Samples of Benthic Invertebrates*, 2nd edn. Freshwater Biological Association, Ambleside, Cumbria, UK.

Ellis, J.C. (1989): *Handbook on the Design and Interpretation of Monitoring Programs*. Water Research Centre. Medmenham.

Ellis, J.C. and Lacy, R.F. (1980): Sampling: defining the task and planning the scheme. *Water Pollut. Control 79:* 452.

El-Shaarawi, A.H. and Esterby, S.R. (1992): Replacement of censored observations by a constant: an evaluation. *Water Res. 26:* 835–844.

Environment Australia (1996): *Australia: State of the Environment.* Environment Australia, Commonwealth Government, Canberra.

Environment Australia (1998): *Environmental Indicators Developed for National State of the Environment Reporting.* Environment Australia, Commonwealth Government, Canberra.

Environment Canada (1990a): *Biological Test Method: Acute Lethality Test Using* Daphnia *spp*. Environmental Protection Series Report EPS 1/RM/11.

Environment Canada (1990b): *Guidance Document on Control of Toxicity Test Precision Using Reference Toxicants*. Report EPA 1/RM/12, Environment Canada, Conservation and Protection, Ottawa, Ontario, Canada.

Environment Canada (1992): Growth inhibition test using the freshwater alga *Selenastrum capricornutum*. Environmental Protection Series Report EPS 1/RM/25.

Environment Canada (1994): *Guidance Document on Collection and Preparation of Sediments for Physicochemical Characterisation and Biological Testing*. Report EPS 1/RM/29.

EPA (1992): *Pre-commissioning Phase Contaminants in Sediments (Sydney Deepwater Outfalls Environmental Monitoring Program)*. Prepared by L. Gray. NSW Environment Protection Authority Marine Section. Report Volume 10, SPCC TR 92/91.

EPAV (1992): *Guidelines for the Design of Water Quality Surveys*. Environmental Protection Authority, Water and Natural Resources Section.

Esterby, S.R. (1996): Review of methods for the detection and estimation of trends with emphasis on water quality applications. *Hydrological Processes 10:* 127–149.

Fairweather, P.G. (1991): Statistical power and design requirements for environmental monitoring. *Australian Journal of Marine and Freshwater Research 42:* 555–567.

Fairweather, P. and Napier, G. (1998): *Environmental Indicators for National State of the Environment Reporting — Inland Waters.* Australia: State of the Environment (Environmental Indicator Reports), Department of the Environment, Canberra.

At URL: http://www.environment.gov.au/soe/envindicators/inland-ind.html

Falconer, I.R. (1994): Comparative evaluation of methods for testing the toxin content of cyanobacterial blooms and related water supplies. *Proc. Int. Workshop on Toxic Cyanobacteria: Current Status of Research and Development*, Adelaide, March 1994.

Fore, L.S., Karr, J.R. and Wisseman, R.W. (1996): Assessing invertebrate responses to human activities: evaluating alternative approaches. *Journal of the North American Benthological Society 15:* 212–231.

Fox, D.R. (1999): Setting water quality guidelines — a statistician's perspective. *SETAC News*, May 1999, 17–18.

Fox, D.R. (2000 unpublished): Using beta-content tolerance intervals to derive water quality guidelines. Available from CSIRO Environmental Projects Office, at URL: http:www.epo.csiro.au/library/wqguideline.pdf

Franklin, N., Stauber, J.L., Markich, S.M. and Lim, R.P. (2000): pH-dependent toxicity of copper and uranium to a tropical freshwater alga (*Chlorella* sp.). *Aquatic Toxicology 48:* 275–289.

Galpin, J.S. and Basson, B. (1990): Some aspects of analysing irregularly spaced time dependent data. *South African Journal of Science 86:* 458–461.

Garner, F.C., Stapanian, M.A. and Fitzgerald, K.E. (1991): Finding causes of outliers in multivariate data. *Journal of Chemometrics 5:* 241–248.

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

Gilliom, R.J. and Helsel, D.R. (1986): Estimation of distributional parameters for censored trace level water quality data. 1: estimation techniques. *Water Resources Research 22:* 135–146.

Gilliom, R.J., Hirsch, R.M. and Gilroy, E.J. (1984): Effect of censoring trace-level water quality data on trend detection capability. *Environmental Science Technology 18:* 530–535.

Gleason, H.A. (1922): On the relation between species and area. *Ecology 3:* 158–162.

Gore, J.A. (Ed.) (1985): *The Restoration of Rivers and Streams*. Butterworth, Stoneham.

Gower, J.C. (1987): Measures of similarity, dissimilarity and distance. In: *Encyclopedia of Statistical Sciences Vol. 5* (S. Kotz and N.L. Johnson, eds), pp. 397–405. Wiley, New York, USA.

Grasshoff, K., Kremling, K. and Ehrhardt, M. (1999): *Methods of Seawater Analysis*. Third Edition, Wiley-VCH, Weinheim.

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

Green, R.H. (1989): Power analysis and practical strategies for environmental monitoring. *Environmental Research 50:* 195–205.

Green, R.H. (1994): Aspects of power analysis in environmental monitoring. In *Statistics in Ecology and Environmental Monitoring* (D.J. Fletcher and B.F.J. Manly, eds), pp. 173–182. University of Otago Press, Dunedin, New Zealand.

Growns, J.E., Chessman, B.C., McEvoy, P.K. and Wright, I.A. (1995): Rapid assessment of rivers using macro-invertebrates: a case study in the Nepean River and Blue Mountains, NSW. *Aust. J. Ecol. 20:* 130–141.

Growns, J.E., Chessman, B.C., Jackson, J.E. and Ross, D.G. (1997): Rapid assessment of Australian rivers using macroinvertebrates: cost and efficiency of six methods of sample processing. *J. North Amer. Benthol. Soc. 16:* 682–693.

Growns, J.E., King, A.J. and Betts, F.M. (1999): The snag bag, a new method for sampling in situ macroinvertebrate communities on large woody debris. *Hydrobiologia 405:* 67–77.

Gunther, A.J., Spies, R.B., Stegeman, J., Woodin, B., Carney, D., Oakden, J. and Hain, L. (1997): EROD activity in fish as an independent measure of contaminant-induced mortality of invertebrates in sediment bioassays. *Marine Environmental Research 44:* 41–49.

Gunthorpe, L., Nottage, M., Palmer, D. and Wu, R. (1995): The development of a fertilisation inhibition bioassay using gametes of the brown alga *Hormosira banksii*. *Australasian Journal of Ecotoxicology 1:* 25–31.

Gunthorpe, L., Nottage, M., Palmer, D. and Wu, R. (1997): *Development and Validation of a Fertilisation Inhibition Bioassay That Uses Gametes of* Hormosira banksii. National Pulp Mills Research Program Technical Report No. 21, Canberra, CSIRO.

Gy (1986): The analytical and economic importance of correctness in sampling. *Anal. Chim. Acta 190:* 13–23.

Hall, G.E.M. (1998): Relative contamination levels observed in different types of bottle used to collect water samples. *Explore 101:* 3–7.

Harris, D.J. and Keffer, W.J. (1974): Wastewater sampling methodologies and flow measurement techniques. US Environmental Protection Agency, Kansas City MO, 1974 EPA 907-974-005.

Harris, J.H. (1995): The use of fish in ecological assessments. *Aust. J. Ecol. 20:* 65–80.

Harris, J.H. and Gehrke, P.C. (1997): *Fish and Rivers in Stress: The NSW Rivers Survey*. NSW Fisheries Office of Conservation, and Cooperative Centre for Freshwater Ecology.

Harris, J.H. and Silveira, R. (1999): Large-scale assessments of river health using an Index of Biotic Integrity with low-diversity fish communities. *Freshwater Biology 41:* 235–252.

Hart, B.T. and Davies, S.H.R. (1977): Batch method for the determination of ion exchangeable trace metals in natural waters. *Aust. J. Mar. Freshwater Res. 28:* 397–402.

Hart, B.T., McKelvie, I.D. and Benson, R. (1993): Real-time instrumentation for monitoring water quality: an Australian perspective. *TrAC 12:* 403–412.

Hart, B.T., Ottaway, E.M. and Noller, B.N. (1987): Magela Creek system Northern Australia I. 1982–83 Wet-season water quality. *Australian Journal of Marine and Freshwater Research 38:* 261–288.

Haski, R., Cardilini, G. and Bartolo, W.C. (1997): *Laboratory Safety Manual: An Essential Manual for Every Laboratory*. CCH Australia Ltd.

Hastie, T.J. and Tibshirani, R.J. (1990): *Generalized Additive Models*. Chapman and Hall.

Hayes, C.R., Warn, A.E. and Green, L.A. (1985): Development of comprehensive water supply quality control in Anglian Water. *J. Inst. Water Eng. Sci. 39:* 539–547.

Hayslip, G.A. (1993): *EPA Region 10 In-stream Biological Monitoring Handbook (for Wadable Streams in the Pacific Northwest)*. US Environmental Protection Agency-Region 10, Environmental Services Division, Seattle, Washington, EPA-910-9-92-013.

Hellawell, J.M. (1986): *Biological Indicators of Freshwater Pollution*. Elsevier, London and New York.

Helsel, D.R. and Gilliom, R.J. (1986): Estimation of distributional parameters for censored trace level water quality data. 2: Verification and applications. *Water Resources Research 22:* 147–155.

Helsel, D. and Hirsch, R.M. (1992): *Statistical Methods in Water Resources*. Elsevier, Amsterdam.

Helsel, D.R. and Hirsch, R.M. (2000): *Statistical Methods in Water Resources*. Studies in Environmental Science 49. Elsevier, New York.

Hickey, C.W. (1988): Benthic chamber for use in rivers: testing against oxygen mass balances. *J. Environmental Engineering 114:* 828–845.

Hill, A.B. (1965): The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine 58:* 295–300.

Hogan, A. (1998): Development and application of marine toxicity identification evaluation (TIE) using unicellular algae. B.Sc. Honours Thesis, University of Technology, Sydney.

Holdway, D.A., Brennan, S.E. and Ahokas, J.T. (1994): Use of hepatic MFO and blood enzyme biomarkers in sand flathead (*Platycephalus bassensis*) as indicators of pollution in Port Phillip Bay, Australia. *Marine Pollution Bulletin 28:* 683–695.

Holdway, D.A., Brennan, S.E. and Ahokas, J.T. (1995): Short review of selected fish biomarkers on xenobiotic exposure with an example using fish hepatic mixed-function oxidase. *Aust. J. Ecol. 20:* 34–44.

Hotzel, G. and Croome, R. (1998): *A Phytoplankton Methods Manual for Australian Rivers*. Occasional Paper No. 18/98. Land and Water Resources Research and Development Corporation, Canberra.

Humphrey, C.L., Bishop, K.A. and Brown, V.M. (1990): Use of biological monitoring in the assessment of effects of mining wastes on aquatic ecosystems of the Alligator Rivers Region, tropical northern Australia. *Environ. Monit. Assess. 14:* 139–181.

Humphrey, C.L., Faith, D.P. and Dostine, P.L. (1995): Baseline requirements for assessment of mining impact using biological monitoring. *Aust. J. Ecol. 20:* 150–166.

Hunt, D.T.E. and Wilson, A.L. (1986): *The Chemical Analysis of Water, General Principles and Techniques*. Royal Society of Chemistry, Cambridge.

Hurlbert, S.H. (1984): Pseudoreplication and the design of ecological field experiments. *Ecological Monographs 54:* 187–211.

Hyne, R.V. and Everett, D.A. (1998): Application of a benthic euryhaline amphipod *Corophium* sp. as a sediment toxicity testing organism for both freshwater and estuarine systems. *Archives of Environmental Contamination and Toxicology 34:* 26–33.

Hyne, R.V. and Wilson, S.P. (1997): Toxicity of acid-sulfate soil leachate and aluminium to the embryos and larvae of Australian bass (*Macquaria novemaculeata*) in estuarine water. *Environ. Pollut. 97:* 221–227.

Hyne, R.V., Rippon, G.D., Hunt, S.M. and Brown, G.H. (1996): Procedures for the Biological Toxicity Testing of Mine Waste Waters Using Freshwater Organisms. Supervising Scientist Report 100, Supervising Scientist, Canberra.

ISO (1989): Water quality — fresh water algal growth inhibition test with *Scenedesmus subspicatus* and *Selenastrum capricornutum*. International Standard ISO 8692: 1989(E).

John, J. (1998): Evaluation of attached diatoms as a tool for riverine bioassessment of water quality. Land and Water Resources Research and Development Corporation Project UCW 3.

Johnson, D.H. (1999): The insignificance of statistical significance testing. *Journal of Wildlife Management 63:* 763–772.

Julli, M., Chapman, J.C. and Thompson, G.B. (1990): Use of Australian cladocerans to generate life-cycle toxicity data. *Environ. Monit. Assess. 14:* 353–362.

Karr, J.R. (1981): Assessment of biotic integrity using fish communities. *Fisheries 6:* 21–27.

Karr, J.R. (1991): Biological integrity: a long-neglected aspect of water resource management. *Ecological Applications 1:* 66–84.

Karr, J.R. and Chu, E.W. (1997): Biological monitoring: essential foundation for ecological risk assessment. *Human and Ecological Risk Assessment 3:* 993–1004.

Karr, J.R. and Chu, E.W. (1999): *Restoring Life in Running Waters: Better Biological Monitoring*. Island Press, Washington, DC, USA.

Karr, J.R. and Dudley, D.R. (1981): Ecological perspective on water quality goals. *Environmental Management 5:* 55–68.

Karr, J.R., Fausch, K.D., Angermeir, P.L., Yant, P.R. and Schlosser, I.J. (1986): Assessing biological integrity in running waters: a method and its rationale. Special Publication No. 5 Illinois Natural History Survey, Champaign Illinois, USA.

Kay, W.R., Smith, M.J., Pinder, A.M., Mcrae, J.M., Davis, J.A. and Halse, S.A. (1999): Patterns of distribution of macroinvertebrate families in rivers of north-western Australia. *Freshwater Biology 41:* 299–316.

Keith, L.H. (1991): *Environmental Sampling and Analysis: A Practical Guide.* Lewis Publishers, Chelsea, Michigan.

Keough, M.J. and Mapstone, B.D. (1995): *Protocols for Designing Marine Ecological Monitoring Programs Associated With BEK Mills.* National Pulp Mills Research Program No. 11, Canberra, CSIRO.

Keough, M.J. and Mapstone, B.D. (1997): Designing environmental monitoring for pulp mills in Australia. *Water Science and Technology 35:* 397–404.

Kerans, B.L. and Karr, J.R. (1994): A benthic index of biotic integrity (B-IBI) for rivers of the Tennessee Valley. *Ecological Applications 4:* 768–785.

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.

Kevekordes, K. and Clayton, M.N. (1996): Using developing embyros of *Hormosira banksii* (Phaeophyta) as a marine bioassay system. *Int. J. Plant Science 157:* 582–585.

Kitchens, L.J. (1987): *Exploring Statistics: A Modern Introduction*. West Publishing Co.

Krassoi, R., Everett, D. and Anderson, I. (1996): *Protocol for Using Doughboy Scallop* Chlamys asperrima *(Mollusca: Pectinidae) L. to Test the Sublethal Toxicity of Single Compounds and Effluents*. National Pulp Mills Research Program Technical Report No. 17, Canberra, CSIRO.

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

Kumar, A. and Chapman, J.C. (1998): Profenofos toxicity to the eastern rainbow fish (*Melanotaenia duboulayi). Environmental Toxicology and Chemistry 17:* 1799–1806.

Lambert, D., Hogg, I. and Maher, W. (1992): Changes in phosphorus fractions during storage of lake water. *Water Research 26:* 645–648.

Legendre, P. and Legendre, L. (1998): *Numerical Ecology.* 2nd edn. Elsevier, Amsterdam.

Lenat, D.R. (1993): A biotic index for the southeastern United States: derivation and list of tolerance values, with criteria for assigning water quality ratings. *Journal of the North American Benthological Society 12:* 279–290.

Lenat, D.R. and Barbour, M.T. (1994): Using benthic macroinvertebrate structure for rapid, cost-effective, water quality monitoring: rapid bioassessment. *Biological Monitoring of Aquatic Ecosystems* (S.L. Loeb and A. Spacie, eds), pp. 187–215. Lewis Publishers, Boca Raton, FL, USA.

Leonard, A., Hyne, R.V., Lim, R.P. and Chapman, J.C. (1999): Effect of endosulfan runoff from cotton fields on macroinvertebrates in the Namoi River. *Ecotoxicol. Environ. Saf. 42:* 125–134.

Leonard, A., Hyne, R.V., Lim, R.P., Leigh, K.A., Jiawei Le and Beckett, R. (in press): Particle fractionation and toxicity of endosulfan contaminated sediment. *J. Environmental Quality*.

Lettenmaier, D.P., Anderson, D.E. and Brenner, R.N. (1984): Consolidation of a stream quality monitoring network. *Water Res. Bull. 20:* 473–481.

Lewis, D.L. (1988): Assessing and controlling sample contamination. *Principles of Environmental Sampling* (I.D. Keith, ed.). ACS Professional Reference Book American Chemical Society, pp. 130–144.

Liu, S., Lu, J., Kolpin, D.W. and Meeker, W.Q. (1997): Analysis of environmental data with censored observations. *Environmental Science Technology 31:* 3358–3362.

Magmann, P. (1991): Unrecognized behaviour and sampling limitations can bias field data. *Environ. Biol. Fish 31:* 403–406.

Maher, W. and Cullen, P. (1997): Design of nutrient programs addressing the Why question. *Proceedings of Workshop on Sampling Nutrients in Aquatic Environments: Collecting Valid and Representative Samples* (K.N. Markwort and W.A. Maher, eds), pp. 2–9. University of Canberra 21–22 April 1997. CRC Freshwater Ecology/ National Eutrophication Management Program.

Maher, W.A., Cullen, P.W. and Norris, R.H. (1994): A framework for designing sampling programs. *Environmental Monitoring and Assessment 30:* 139–162.

Maher, W.A. and Woo, L. (1998): Procedures for the storage and digestion of natural waters for the determination of filterable reactive phosphorus, total filterable phosphorus and total phosphorus. *Anal. Chim. Acta 375:* 5–47.

Mapstone, B.D. (1995): Scalable decision rules for environmental impact studies: effect size, Type I, and Type II Errors. *Ecological Applications 5:* 401–410.

Marchant, R., Hirst, A., Norris, R. and Metzeling, L. (1999): Classification of macroinvertebrate communities across drainage basins in Victoria, Australia: consequences of sampling on a broad spatial scale for predictive modelling. *Freshwater Biology 41:* 253–268.

Margalef, D.R. (1958): Information theory in ecology. *Gen. Syst. 3:* 36–71.

Markich, S.J. and Camilleri, C. (1997): Investigation of metal toxicity to tropical biota: recommendations for revision of the Australian water quality guidelines. Supervising Scientist report 127, Supervising Scientist, Canberra.

McBride, G.B., Loftis, J.C. and Adkins, N.C. (1993): What do significance tests really tell us about the environment? *Environmental Management 17:* 423–432.

McCullagh, P. and Nelder, J.A. (1983): *Generalized Linear Models*. Chapman and Hall.

McCullough, R.D. and Wilson, B. (1999): On the accuracy of statistical procedures in Microsoft Excel 97. *Computational Statistics and Data Analysis 31:* 21–37.

McLaren, J.W., Mykytiuk, A.P., Willie, S.N. and Berman, S.S. (1985): Determination of trace metals in seawater by inductively coupled plasma mass spectrometry with preconcentration on silica-immobilised 8-hydroxyquinoline. *Anal. Chem. 57:* 2907–2911.

McNeil, A.R. (1985): *Microbiological Water Quality Criteria: A Review For Australia*. Australian Water Resources Council Technical Paper. Number 85. Australian Government Publishing Service. Canberra.

Millard, S.P., Yearsley, J.R. and Lettenmaier, D.P. (1985): Space–time correlation and its effects on methods for detecting aquatic ecological change. *Canadian Journal of Fisheries and Aquatic Sciences 42:* 1391–1400.

Miller, D.G. and Ellis, J.C. (1986): Derivation and monitoring of consents. *Water Pollution Control* 249–258.

Miller, D.L., Leonard, P.M., Hughes, R.M., Karr, J.R., Moyle, P.B., Schrader, L.H., Thompson, B.A., Daniels, R.A., Fausch, K.D., Fitzhugh, G.A., Gammon, J.R., Halliwell, D.B., Angermeier, P.L. and Orth, D.J. (1988): Regional applications of an index of biotic integrity for water resource management. *Fisheries 13:* 12–20.

Miller, G.A., Galanter, A.E. and Pilbram, K.H. (1960): *Plans and Structure of Behaviour*. Holt, Rinehart and Winston, New York.

Montgomery, D.C. (1985): *Introduction to Statistical Quality Control*. John Wiley and Sons.

Montgomery, H.A.C. and Hart, I.C. (1974): The design of sampling programs 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. Oceanogr. 32:* 1342–1352.

Mudroch, A. and Azcue, J.M. (1995): *Manual of Aquatic Sediment Sampling*. Lewis Publishers, Boca Raton, FL.

Mudroch, A. and Macknight, S.D. (1991): *Handbook of Techniques for Aquatic Sediments Sampling.* CRC Press, Boca Raton, Florida, USA.

Munday, B.L., Nowak, B.F. and Deavin, J.G. (1991): Toxicity of eucalypt-based pulp mill effluents as assessed by acute and chronic bioassays. In: *Environmental Fate and Effects of Bleached Pulp Mill Effluents* (A. Sodergren, ed.). Swedish Environment Protection Agency Report 4031.

Muschal, M. (1998): Central and North West Regions Water Quality Program: 1997/98 Report on Pesticides Monitoring. CNR98.038, Dept of Land and Water Conservation, Centre for Natural Resources, Parramatta.

National Research Council (1987): Committee on Biological Markers. *Environmental Health Perspectives 754:* 3–9.

Neave, H.R. (1988): *Distribution-free Tests.* Unwin Hyman, London.

Neter, J., Kutner, M.H., Nachtsheim, C.J. and Wasserman, W. (1996): *Applied Linear Statistical Models*, 4th edition. Irwin, Illinois.

Nolting, R.F. and de Jong, J.T.M. (1994): Sampling and analytical methods for the determination of trace metals in seawater. *Intern. J. Environ. Anal. Chem. 57:* 189–196.

Norris, R.H. (1986): Mine waste pollution of the Molonglo River, New South Wales and the Australian Capital Territory: effectiveness of remedial works at Captains Flat mining area. *Aust. J. Mar. Freshwater Res. 37:* 147–157.

Norris, R.H. and Georges, A. (1986): Design and analysis for assessment of water quality. In *Limnology in Australia*. (P. De Deckker and W.D. Williams, eds), pp. 555–572. CSIRO, Melbourne and W. Junk, Dordrecht.

Norris, R.H. and Thoms, M.C. (1999): What is river health? *Freshwater Biology 41:* 197–209.

Norris, R.H., McElravy, E.P. and Resh, V.H. (1992): The sampling problem. In *Rivers Handbook* (P. Calow and G.E. Petts, eds), pp. 1–15. Blackwell Scientific Publishers, Oxford, England.

Nriagu, J.O., Lawson, G., Wong, H.K.T. and Azcue, J.M. (1993). A protocol for minimising contamination in the analysis of trace metals in Great Lakes waters. *J. Great Lakes Res. 19:* 175–182.

Odum, H.T. (1956): Primary production in flowing waters. *Limnol. Oceanogr. 1:* 102–117.

OECD (1981): *Guidelines for Testing of Chemicals.* Organisation for Economic Cooperation and Development, Paris.

OECD (1984): *OECD Guidelines for Testing of Chemicals.* Vol. 1. Organisation for Economic Cooperation and Development, Paris.

Osenberg, C.W. and Schmitt, R.J. (1994): Detecting human impacts in marine habitats. *Ecological Applications 4:* 1–2.

Ott, L. (1984): *An Introduction to Statistical Methods and Data Analysis.* Duxbury Press, Boston.

Pablo, F., Chapman, J.C., Julli, M. and Patra, R. (1997): Toxicity characterisation of sewage treatment plant and creek water samples in the Sydney district using TIE techniques. *Proceedings of 4th Annual Conference.* Australasian Society for Ecotoxicology, Brisbane, July 1997, abstract O9.2.

Parkhurst, D.F. (1998): Arithmetic versus geometric means for environmental concentration data. *Environmental Science and Technology News.* February 1, 1998. American Chemical Society.

Parry, G.D., Hobday, D.R., Currie, D.R. and Officer, R.A. (1996): *The Distribution, Abundance and Diets of Demersal Fish in Port Phillip Bay*. Technical Report 21, Port Phillip Bay Environmental Study, CSIRO Publishing, Collingwood, Australia.

Parsons, T.R., Maita, Y. and Lalli, C. (1985): *A Manual of Chemical and Biological Methods for Seawater Analysis.* Pergamon.

Peterson, S.M. and Stauber, J.L. (1996): New algal enzyme bioassay for the rapid assessment of aquatic toxicity. *Bull. Environ. Contam. Toxicol. 56:* 750–757.

Peterson, S.M., Apte, S.C., Batley, G.E. and Coade, G. (1995): Passive sampler for chlorinated pesticides in estuarine waters. *Chemical Speciation and Bioavailability 7:* 83–88.

Plafkin, J.L., Barbour, M.T., Porter, K.D., Gross, S.S. and Hughes, R.M. (1989): *Rapid Bioassessment Protocols for Use in Streams and Rivers. Benthic Macroinvertebrates and Fish*. EPA/440/4-89/001, Office of Water Regulations and Standards, US Environmental Protection Agency, Washington, DC, USA.

Please, P.M., Watkins, K.L. and Bauld, J. (1997): *A Groundwater Quality Assessment of the Alluvial Aquifer in the Logan–Albert Catchment, South-East Queensland*. Australian Geological Survey Organisation.

Pratt, J.W. (1976): A discussion of the question: for what use are tests of hypotheses and tests of significance. *Communications in Statistics, Series A5:* 779–787.

QDME (1995): *Technical Guidelines for the Environmental Management of Exploration and Mining in Queensland*. Queensland Department of Mines and Energy.

Radford, P.J. and West, J. (1986): Models to minimize monitoring. *Water Res. 20:* 1059–1066.

Rapport, D.J. (1991): Myths in the foundation of economics and ecology. *Biological Journal of the Linnaean Society 44:* 185–202.

Rayment, G.E. and Higginson, R. (1992): *Australian Laboratory Handbook of Soil and Water Chemical Methods*. Inkata Press, Sydney.

RD Instruments (1989): *Acoustic Doppler Current Profilers: Principles of Operation: A Practical Primer*. RD Instruments, San Diego.

Reimann, C., Siewers, U., Skarphagen, H. and Banks, D. (1999): Does bottle type and acid-washing influence trace element analyses by ICP-MS on water samples? *Sci Tot. Environ. 239:* 111–130.

Resh, V.H. and Jackson, J.K. (1993): Rapid assessment approaches to biomonitoring using benthic macroinvertebrates. In *Freshwater Biomonitoring and Benthic Macroinvertebrates* (D.M. Rosenberg and V.H. Resh, eds), pp. 195–233. Chapman and Hall, New York.

Resh, V.H., Norris, R.H. and Barbour, M.T. (1995): Design and implementation of rapid assessment approaches for water resource monitoring using benthic macroinvertebrates. *Australian Journal of Ecology 20:* 108–121.

Robson, A.J. and Neal, C. (1996): Water quality trends at an upland site in Wales, UK, 1983–1993. *Hydrological Processes 10:* 183–203.

Rousseeuw, P.J. and Yohai, V. (1984): Robust regression by means of *S*-estimators. In *Robust and Nonlinear Time Series Analysis.* (J. Franke, W. Hardle and R.D. Martin, eds), pp. 256–272. Lecture Notes in Statistics 26. Springer-Verlag.

Rossi, R.E., Mulla, D.J., Journel, A.G. and Franz, E.H. (1992): Geostatistical tools for modeling and interpreting ecological spatial dependence. *Ecological Monographs 62:* 277–314.

Schneider, P. and Wyllie, S.J. (1991): An efficient vibrocoring system for collecting coastal sediments: a comparison with other techniques. In *Coastal Engineering — Climate for Change*. Proceedings 10th Australasian Conference on Coastal and Ocean Engineering Auckland 2–6 December 1991, 357–362.

Schofield, N.J. and Davies, P.E. (1996): Measuring the health of our rivers. *Water: Journal of the Australian Water and Wastewater Association 23:* 39.

Segar, D.A., Phillips, D.J.H. and Stamman, E. (1987): Strategies for long term pollution monitoring of the Coastal Oceans. In *New Approaches to Monitoring Aquatic Ecosystems* (T.P. Boyle, ed.), pp. 12–27. ASTM STP 940. American Society for Testing and Materials , Philadelphia, USA.

Shackleford, B. (1988): *Rapid Bioassessments of Lotic Macroinvertebrate Communities: Biocriteria Development.* Arkansas Department of Pollution Control and Ecology, Little Rock, Arkansas.

Shannon, C.E. (1948): A mathematical theory of communication. *Bell System Tech. J. 27:* 379–423, 623–656.

Sharp, W.E. (1971): A topologically optimum water-sampling plan for rivers and streams. *Water Resources Res. 7:* 1641–1646.

Short, C.S. (1980): Sampling programme design for water quality in distribution. *Water Serv. 84:* 529.

Simon, J. and Laginestra, E. (1997): *Bioassay for Testing Sublethal Toxicity in Effluents Using Gametes of Sea Urchin* Heliocidaris tuberculata. National Pulp Mills Research Program Technical Report No. 20, Canberra, CSIRO.

Simpson, E.H. (1949): Measurement of diversity. *Nature 163:* 688.

Smith, E.P. and Voshell, J.R. Jr (1997): *Studies of Benthic Macroinvertebrates and Fish in Streams within EPA Region 3 for Development of Biological Indicators of Ecological Condition*. Virginia Polytechnic Institute and State University, Blacksburg, VA, USA.

Sokal, R.R. and Rohlf, F.J. (1995): *Biometry: The Principles and Practice of Statistics in Biological Research*, 3rd edn. Freeman, New York.

Spiegel, M.R. (1992): *Theory and Problems of Statistics*. McGraw-Hill, London.

Stark, J.D. (1985): *A Macroinvertebrate Community Index of Water Quality for Stony Streams*. Water and soil miscellaneous publication 87, Wellington, NZ.

Stark, J.D. (1993): Performance of the Macroinvertebrate Community Index: effects of sampling method, sample replication, water depth, current velocity and substratum on index values. *New Zealand Journal of Marine and Freshwater Research* 27: 463–478.

Stark, J.D. (1998): The SQMCI: a biotic index for freshwater macroinvertebrate coded abundance data. *New Zealand Journal of Marine and Freshwater Research 32:* 55–66.

State of the Environment Advisory Council (Eds) (1996): *Australia: State of the Environment 1996*, an independent report to the Commonwealth Minister for the Environment. Department of the Environment, Sport and Territories, (CSIRO Publishing).
At URL: http://www.environment.gov.au/soe/soe96/index.html

Stauber, J.L. (1995): Toxicity testing using marine and freshwater unicellular algae. *Australasian J. Ecotox. 1:* 15–24.

Stauber, J.L., Tsai, J., Vaughan, G.T., Peterson, S.M. and Brockbank, C.I. (1994): *Algae as Indicators of Toxicity of the Effluent from Bleached Eucalypt Kraft Pulp Mills*. National Pulp Mills Research Program Technical Report No. 3, Canberra, CSIRO.

Stauber, J.L., Ahsanullah, M., Nowak, B., Eriksen, R. and Florence, T.M. (1996): Toxicity assessment of waters from Macquarie Harbour, western Tasmania, using algae, invertebrates and fish. Mount Lyell Remediation Research and Demonstration Program, Supervising Scientist Report 112, ACT.

Stevens, A., Devlin, M., Brodie, J., Baer, M. and Loury, M. (1996): Spatial influence and composition of river plumes in the Central Great Barrier Reef. In *Downstream Effects of Landuse*. (H.G. Hunter, A.G. Eyles and G.E. Rayment, eds), pp. 85–92. Department of Natural Resources, Queensland, Australia.

Stevens, S.S. (1946): On the theory of scales of measurement. *Science 103:* 677–680.

Stewart-Oaten, A., Murdoch, W.W. and Parker, K.R. (1986): Environmental impact assessment: 'pseudoreplication' in time? *Ecology 67:* 929–940.

Suter, G.W. (1993): Retrospective risk assessment. *Ecological Risk Assessment* (G.W. Suter, L.W. Barnthouse, S.M. Bartell *et al.*, eds), Lewis Publishers, Ann Arbor, Michigan, USA.

Thomas, L. and Krebs, C.J. (1997): A review of statistical power analysis software. *Bulletin of the Ecological Society of America 78*: 129–139.

Thompson, K.W., Deaton, M.L., Foutz, R.V., Cairns, J. Jr and Hendricks, A.C. (1982): Application of time-series intervention analysis to fish ventilatory response data. *Can. J. Fish. Aquat. Sci. 39:* 518–521.

Thompson, S.K. (1992): *Sampling*. Wiley, New York.

Tufte, E.R. (1983): *The Visual Display of Quantitative Information*. Graphics Press, Cheshire Co.

Turak, E., Flack, L.K., Norris, R.H., Simpson, J. and Waddell, N. (1999): Assessment of river condition at a large spatial scale using predictive models. *Freshwater Biology 41:* 283–298.

Twining, J.R. and Nowak, B. (1996): Acute toxicity of a sulfide-bearing marine effluent to Australian bass and blue mussels. *Australasian J. Ecotox. 2:* 67–74.

Underwood, A.J. (1990): Experiments in ecology and management: their logics, functions and interpretations. *Human and Ecological Risk Assessment 2:* 331–347.

Underwood, A.J. (1991): Beyond BACI: experimental designs for detecting human environmental impacts on temporal variation in natural populations. *Australian Journal of Marine and Freshwater Research 42:* 569–587.

Underwood, A.J. (1992): Beyond BACI: the detection of environmental impact on populations in the real, but variable, world. *Journal of Experimental Marine Biology and Ecology 161:* 145–178.

Underwood, A.J. (1994): On beyond BACI: sampling designs that might reliably detect environmental differences. *Ecological Applications 4:* 3–15.

Underwood, A.J. (1996): *Environmental Design and Analysis in Marine Environmental Sampling*. IOC Manuals and Guides No. 34. UNESCO.

USEPA (1982): *Handbook for Sampling and Sample Preservation of Water and Wastewater*. US Department of Commerce National Technical Information Service, Cincinatti, Ohio, USA.

USEPA (1983): *Methods for Chemical Analysis of Water and Wastes*. EPA/600/4-79-020, Revised March 1983. Environmental Monitoring Systems Support Laboratory, Office of Research and Development, US Environmental Protection Agency, Cincinnati, Ohio, USA.

USEPA (1991a): *Methods for Aquatic Toxicity Identification Evaluations. Phase I Toxicity Characterisation Procedures*. 2nd edn. EPA-600/6-91/0303. Office of Research and Development, Duluth, MN, USA.

USEPA (1991b): *Evaluation of Dredged Material Proposed for Ocean Disposal. Testing Manual*. US Environmental Protection Agency/US Army Corps of Engineers, Report EPA-503/8-91/001, Washington DC, USA.

USEPA (1993): *Methods for Measuring the Acute Toxicity of Effluents and Receiving Waters to Freshwater and Marine Organisms*. 4th edn. United States Environmental Protection Agency EPA/600/4-90/027. Office of Research and Development, Cincinnati, Ohio, USA.

USEPA (1994a): *Short-Term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Water to Marine and Estuarine Organisms*. 2nd edn. United States Environmental Protection Agency EPA-600-4-91-003. Office of Research and Development, Cincinnati, Ohio, USA.

USEPA (1994b): *Short-Term Methods for Estimating the Chronic Toxicity of Effluents and Receiving Water to Freshwater Organisms*. United States Environmental Protection Agency EPA-600-4-91-002. Office of Research and Development, Cincinnati, Ohio, USA.

USEPA (1994c): *Methods for the Determination of Metals in Environmental Samples — Supplement 1*. Report No. EPA/600/R-94-111. United States Environmental Protection Agency, Cincinnati, Ohio.

USEPA (1995): Code of Federal Regulations (CFR) 40, Part 136. *Methods for Organic Chemicals Analysis of Municipal and Industrial Wastewater*. US Environmental Protection Agency, Washington DC, USA.

USEPA (1996a): *Marine Toxicity Identification Evaluation (TIE) Phase I Guidance Document*. EPA/600/R-96/054. Office of Research and Development, Washington, DC, USA.

USEPA (1996b): *Compendium of ERT Surface Water and Sediment Sampling Procedures*. Report No. EPA 540/P-91-005, US Environmental Protection Agency, Washington, DC.

USEPA (1996c): *Mercury in Water by Oxidation, Purge and Traps and Cold Vapour Atomic Fluorescence Spectrometry*. Method 1631. USEPA Report No. EPA/821/R-96-001. US Environmental Protection Agency, Washington DC, USA.

USEPA (1996d): *Determination of Inorganic Arsenic in Water by Hydride Generation Atomic Absorption Spectrometry*. Method 1632. USEPA Report No. EPA/821/R-96-002. US Environmental Protection Agency, Washington DC, USA.

USEPA (1996e): *Test Methods for Evaluating Solid Waste. Physical/Chemical Methods*. Report No. 530/SW-846. US Environmental Protection Agency, Washington DC, USA.

Viarengo, A., Bettella, E., Fabbri, R., Burlando, B. and Lafaurie, M. (1997): Heavy metal inhibition of EROD activity in liver microsomes from the Bass *Dicentrarchus labrax* exposed to organic xenobiotics: role of GSH in the reduction of heavy metal effects. *Mar. Environ. Res. 44:* 1–11.

Wallace, J.B., Webster, J.R. and Woodall, W.R. (1977): The role of filter feeders in flowing waters. *Arch. Hydrobiol. 79:* 506–532.

Walpole, R.E and Myers, R.H. (1985): *Probability and Statistics for Engineers and Scientists.* Macmillan Publishing, 3rd edn. New York.

Ward, A.K. and Johnson, M.D. (1996): Heterotrophic microorganisms. In *Methods in Stream Ecology* (F.R. Hauer and G.A. Lambert, eds), 232–235. Academic Press, San Diego.

Ward, R.C., Loftis, J.C. and McBride, G.B. (1990): *Design of Water Quality Monitoring Systems*. Van Nostrand Reinhold, New York.

Ward, T., Butler, E. and Hill, B. (1998): *Environmental Indicators for National State of the Environment Reporting — Estuaries and the Sea.* Australia: State of the Environment (Environmental Indicator Reports), Department of the Environment, Canberra.

At URL: http://www.environment.gov.au/soe/envindicators/estuaries-ind.html

Warwick, R.M. and Clarke, K.R. (1993): Comparing the severity of disturbance: a meta-analysis of marine macrobenthic community data. *Marine Ecology Progress Series 92:* 221–231.

Warwick, W.F. (1989): Morphological deformities in larvae of *Procladius* Skuse (Diptera: Chironomidae) and their biomonitoring potential. *Can. J. Fish Aquatic Sci. 46:* 1255–1270.

Welsh, D.R. and Stewart, D.B. (1989): Applications of intervention analysis to model the impact of drought and bushfires on water quality. *Australian Journal of Marine and Freshwater Research 40:* 241–257.

Whitton, B.A. and Kelly, M.G. (1995): Use of algae and other plants for monitoring rivers. *Aust. J. Ecol. 20:* 45–56.

WMO (1988): *Manual on Water Quality Monitoring — Planning and Implementation of Sampling and Field Testing*. Operational Hydrology Report No. 27. Secretariat of the World Meteorological Organization, Geneva, Switzerland.

World Health Organization (1993): *Guidelines For Drinking Water Quality, Volume 1: Recommendations*, 2nd edn. World Health Organization, Geneva.

Wright, J.F. (1995): Development and use of a system for predicting macroinvertebrates in flowing waters. *Aust. J. Ecol. 20:* 181–197.

Yohai, V.J. and Zamar, R.H. (1998): Optimally locally robust M-estimates of regression. *Journal of Statistical Inference and Planning.*

Yohai, V., Stahel, W.A. and Zamar, R.H. (1991): A procedure for robust estimation and inference in linear regression. In *Directions in Robust Statistics and Diagnostics, Part III.* (W.A. Stahel and S.W. Weisberg, eds). Springer-Verlag.

Zar, J.H. (1984): *Biostatistical Analysis*. Prentice-Hall, Englewood Clifs, NJ.

Zhang, H. and Davison, W. (1995): Performance characteristics of diffusion gradients in thin films for the in situ measurement of trace metals in aqueous solution. *Anal. Chem. 67:* 3391–3400.

Further reading for Case Studies, Appendix 4:

A4.1. Case Study 1, Burrinjuck Dam and Murrumbidgee River

ACT and Sub-region Planning Committee & Upper Murrumbidgee Catchment Coordination Committee (1994): *Regional Water Quality Study*.

ACT Commissioner for the Environment, State of the Environment Reports 1994, 1995, 1998.

CRC for Freshwater Ecology & CSIRO (1999): Physical and nutrient factors controlling algal succession and biomass in Burrinjuck Reservoir; Milestone Report, National Eutrophication Management Program.

Department of Land & Water Conservation (1997): Water quality, hydrology and fisheries information for water users in the Burrinjuck Sub-catchment, Technical Report No. 97/10.

Environment ACT and Planning and Land Management, ACT Program for the Monitoring River Health Initiative, Reports June 1997, November 1997, May 1998, August 1998.

Environment ACT and Planning and Land Management, Canberra Water Quality Reports 1996, 1997, 1998.

A4.3. Case Study 3, Derwent Estuary

Bloom, H. (1975): *Heavy Metals in the Derwent Estuary*. Department of Chemistry, University of Tasmania.

Davies, P.E. and Kalish, S. (1994): Influence of river hydrology on the dynamics and water quality of the upper Derwent Estuary, Tasmania. *Aust. J. Mar. Freshwater Res. 45:* 109–130.

Department of Environment and Land Management (1995): *The State of the Derwent.* Proceedings of a Scientific Forum.

Department of Environment and Land Management (1996): *State of the Environment Tasmania Volume 1: Conditions and Trends.* State of the Environment Unit, Land Information Services, DELM, Hobart.

Pasminco EZ (1996): *Environmental Management Plan*. John Miedecke & Partners Pty Ltd.

Pasminco EZ (1996a): *Effluent Monitoring Data*. Report to Department of Environment and Land Management.

Pirzl, H.R. (1996): Distributions and Changes of Heavy Metal Concentrations in Sediments of the Derwent Estuary. B.Sc. (Hons) Thesis, Department of Geography and Environmental Studies, University of Tasmania, Hobart.

A4.4. Case Study 4, Great Barrier Reef

Brodie, J. and Furnas, M. (1992): Long term monitoring program for eutrophication and the design of a monitoring program for the Great Barrier Reef. In *Proceedings of the Seventh International Coral Reef Symposium*, Guam 1992. Vol. 1.

Devlin, M., Haynes, D. and Brodie, J. (1997): Cyclone Justin (March 1997) flood plume sampling: preliminary results. In *The Proceedings of the Australian Coral Reef Society 75th Anniversary Conference*, Heron Island 1997, pp. 56–66. School of Marine Science, The University of Queensland, Brisbane.

Greenberg, A.E., Clesceri, L.S. and Eaton, A.D. (1996): *Standard Methods for the Examination of Water and Wastewater*. 18th edition.

Haynes, D., Brodie, J. and Devlin, M. (1997): Long-term monitoring of chlorophyll concentrations Great Barrier Reef World Heritage Area. In *Proceedings of Workshop on Sampling Nutrients in Aquatic Ecosystems: Collecting Valid and Representative Samples*. University of Canberra.

Steven, A.D.L., Trott, L., Pantus, F. and Brooks. D. (1998): *Long-term Chlorophyll Monitoring in the Great Barrier Reef Lagoon: Status Report 1, 1993–1995*. Research Publication Number 55. Great Barrier Reef Marine Park Authority.

Index