GUIDELINES FOR MONITORING BENTHOS IN FRESHWATER ENVIRONMENTS

JANUARY 1993

Environment Canada Environnement Canada

GUIDELINES FOR MONITORING BENTHOS IN FRESHWATER ENVIRONMENTS

Prepared for: **ENVIRONMENT CANADA**
5th Floor, 224 West Esplanade
North Vancouver, B.C.
V7M 3H7

Prepared by: **EVS CONSULTANTS**
195 Pemberton Avenue
North Vancouver, B.C.
V7P 2R4

January 1993

TABLE OF CONTENTS

LIST OF TABLES

Page No.

LIST OF FIGURES

ACKNOWLEDGEMENTS

The guideline document was written by Mr. Wade Gibbons, Dr. Mark Munn and Dr. Michael Paine and reviewed by Dr. Peter Chapman and Mr. Clark Gunter. The document was prepared by Mrs. Arlene Booth, Ms. Jackie Gelling, Ms. Leanne Cook, and Mrs. Elisa Eichen. The figures were produced by Ms. Angela Crampton.

Speakers and other attendees at the Benthos Monitoring Workshop held February, 1992, in Vancouver provided much of the original material, and were all given the opportunity to review a draft. We would like to thank them for their input and comments. We would also like to thank the Scientific Authorities (Mr. Brian Kelso and Mr. George Derksen of Environment Canada) for initiating the Workshop and development of a guidelines document, and for providing support, comments and references.

Citation:

Gibbons, W.N., M.D. Munn and M.D. Paine. 1993. Guidelines for monitoring benthos in freshwater environments. Report prepared for Environment Canada, North Vancouver, B.C. by EVS Consultants, North Vancouver, B.C. 81 pp.

INTRODUCTION

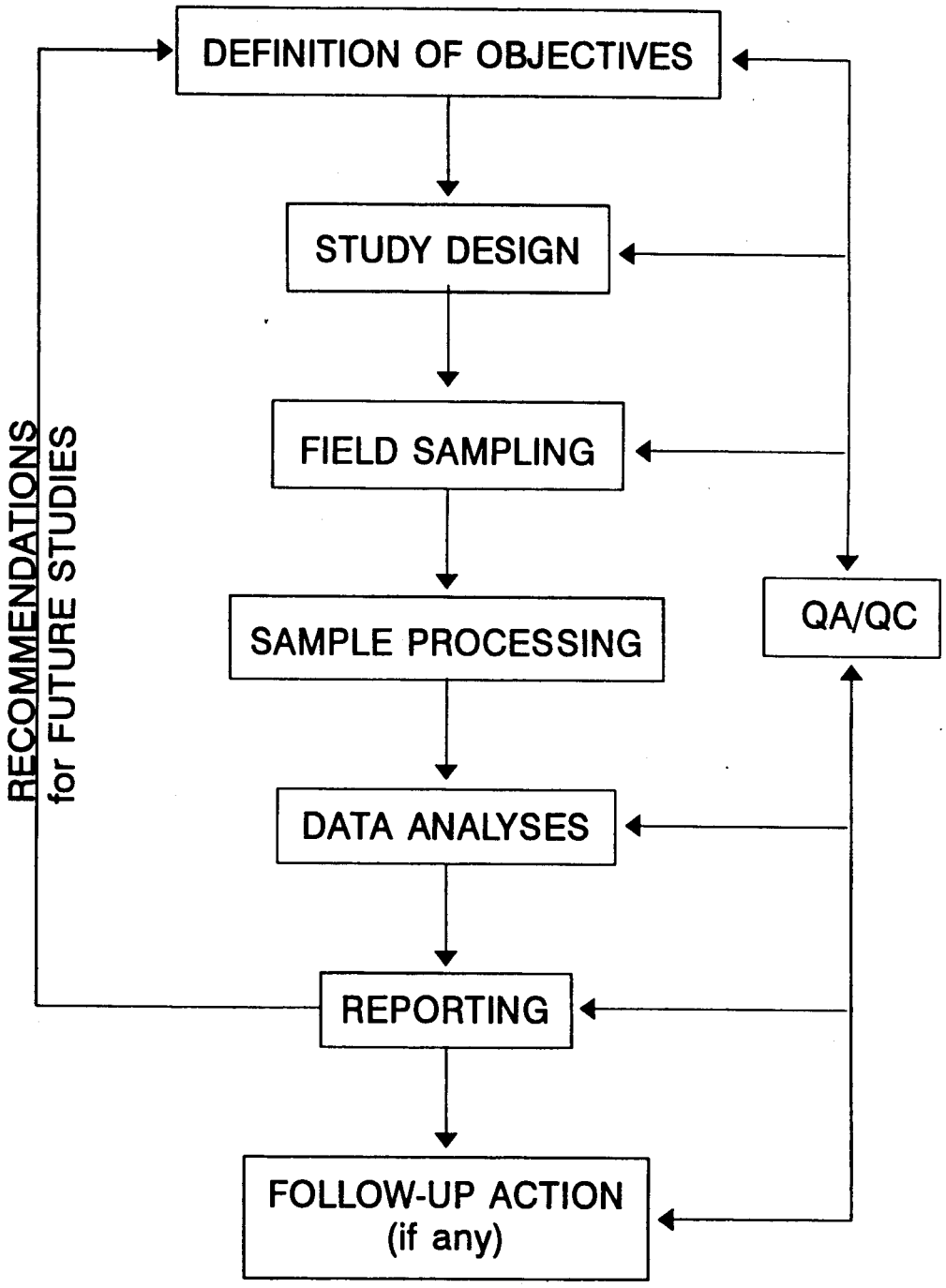

Figure 1. The key elements of a benthos monitoring study.

1.0 INTRODUCTION

1.1 Background

Historically, benthic invertebrates have been viewed as useful organisms for evaluating environmental impacts on aquatic systems (freshwater and marine) (Klemm et al., 1990; Rosenberg and Resh, 1993). They are relatively sedentary organisms, and are sensitive to changes in sediment and water quality. Benthic communities also reflect the cumulative effects of present and past conditions, because they have low mobility and life cycles of several weeks to years (Wilhm, 1975). Their ecological relationships are relatively well understood (Herricks and Cairns, 1982), and they are the major food source for many fish species. For these reasons, sampling benthic communities is regarded as a cost-effective means of assessing the aquatic environment.

Although benthos monitoring programs have been conducted for decades (Cairns and Pratt, 1993), in Canada there has been little effort to standardize the wide array of methods and approaches used. As a first step in achieving standardization, Environment Canada and EVS Consultants hosted a technically-based workshop on benthos monitoring (Gibbons and Booth, 1992) to attempt to develop a consensus on the approach to be used when undertaking benthic invertebrate studies (as an environmental monitoring tool) in freshwater environments. The workshop provided an excellent forum to solicit expert advice and opinion on all aspects of benthos monitoring. Information gathered during the workshop was used to produce a draft guideline document. Workshop attendees reviewed the draft document, and their comments were incorporated in this final comment. The final document represents the first stage in the development of a protocol for freshwater benthos monitoring for the purpose of environmental assessment.

This document emphasizes the ideas, opinions and consensus developed during the workshop discussions. Where there was no consensus, or where topics were not addressed during the workshop, suggestions from the literature have been included for consideration. Readers should also refer to general sources such as Hynes (1960, 1970), Klemm et al. (1990), and Rosenberg and Resh (1993), and specific sources cited in various sections of this report, for more detailed information.

1.2 Definitions and Scope

The U.S. National Research Council has defined monitoring as "the range of activities needed to provide management information about environmental conditions or contaminants" (NRC, 1990, p.7). Activities could include field sampling programs, toxicity tests, chemical analyses, and mathematical models. Rosenberg and Resh (1993) use a similar broad definition of monitoring, but restrict their treatment to biological (as opposed to chemical) monitoring and benthic invertebrates. Several authors in Goldsmith (1991) further restrict monitoring to field surveys of benthic invertebrates (or other organisms), usually as a means of assessing environmental quality over space or time. These restrictive definitions correspond to surveillance as defined by Rosenberg and Resh (1993) and to the scope of this guideline document. This document also emphasizes the assessment of impacts from specific sources, because it was based in part on guidelines developed for the assessment of the effects of pulp and paper mill discharges (Environment Canada, 1991).

The term "macroinvertebrates" is not used in this document, as conflicting definitions based on mesh size of sampling devices or sieves exist. Instead, the general terms "invertebrates" or "benthic invertebrates" are used. We use the term "communities", following general usage, but recognize that "assemblages" may be as appropriate or more appropriate. The scope of this document is also confined to monitoring benthic invertebrates only, but comprehensive and effective monitoring programs also include other organisms such as fish or algae. Any investigator planning a benthos monitoring program should consider whether benthos monitoring is both necessary (i.e., capable of achieving objectives) and sufficient (i.e., capable of achieving objectives even if other organisms are not monitored).

1.3 Overview of Monitoring Components

This guideline document focuses on the following key elements:

- Quality Assurance/Quality Control (QA/QC)
- Definition of Objectives
- Study Design
- Field Sampling
- Sample Processing
- Data Analyses
- Reporting

Although discussed separately, each element must be considered in the context of the remaining elements to develop an integrated program. Any benthos monitoring study must comprise a logical sequence of design events, typically (from Green, 1979):

Purpose 6 Question (objective) 6 Hypotheses 6 Sampling Design 6 Statistical Analysis 6Test of Hypotheses 6 Interpretation/Presentation of Results.

Figure 1 presents an overview of the progressive sequence of study events of a complete benthos monitoring program incorporating the above design components. This overview is followed in this guideline document.

1.4 References Cited

Cairns, J. Jr. and J.R. Pratt. 1993. A history of biological monitoring using benthic macroinvertebrates. pp. 10-27. In: Freshwater Biomonitoring and Benthic Macroinvertebrates. D.M. Rosenberg and V.H. Resh (eds). Chapman and Hall, New York, NY.

Gibbons, W.N. and A.S. Booth (eds). 1992. Proceedings of the benthos monitoring workshop, February 11-12, 1992, Vancouver, B.C. Report prepared for Environment Canada, North Vancouver, B.C. by EVS Consultants, North Vancouver, B.C. 59 pp. + App.

Goldsmith, F.B. 1991. Monitoring for Conservation and Ecology. Chapman and Hall, London, UK.

Green, R.H. 1979. Sampling Design and Statistical Methods for Environmental Biologists. John Wiley and Sons, Toronto, Ont. 257 pp.

Herricks, E.E. and J. Cairns, Jr. 1982. Biological monitoring. Part III - Receiving system methodology based on community structure. Wat. Resour. 16:141-153.

Hynes, H.B.N. 1960. The Biology of Polluted Waters. University of Toronto Press, Toronto, Ont. 202 pp.

Hynes, H.B.N. 1970. The Ecology of Running Waters. University of Toronto Press, Toronto, Ont. 555 pp.

Klemm, D.J., P.A. Lewis, F. Fulk and J.M. Lazorchak. 1990. Macroinvertebrate field and laboratory methods for evaluating the biological integrity of surface waters. EPA 600/4-90/030. U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Cincinnati, OH.

Rosenberg, D.M. and V.H. Resh (eds). 1993. Freshwater Biomonitoring and Benthic Macroinvertebrates. Chapman and Hall, New York, NY. 488 pp.

Wilhm, J.L. 1975. Biological indicators of pollution. pp. 375-402. In: River Ecology. B.A. Whitton (ed.). University of California Press, Berkley, CA.

QUALITY ASSURANCE/QUALITY CONTROL

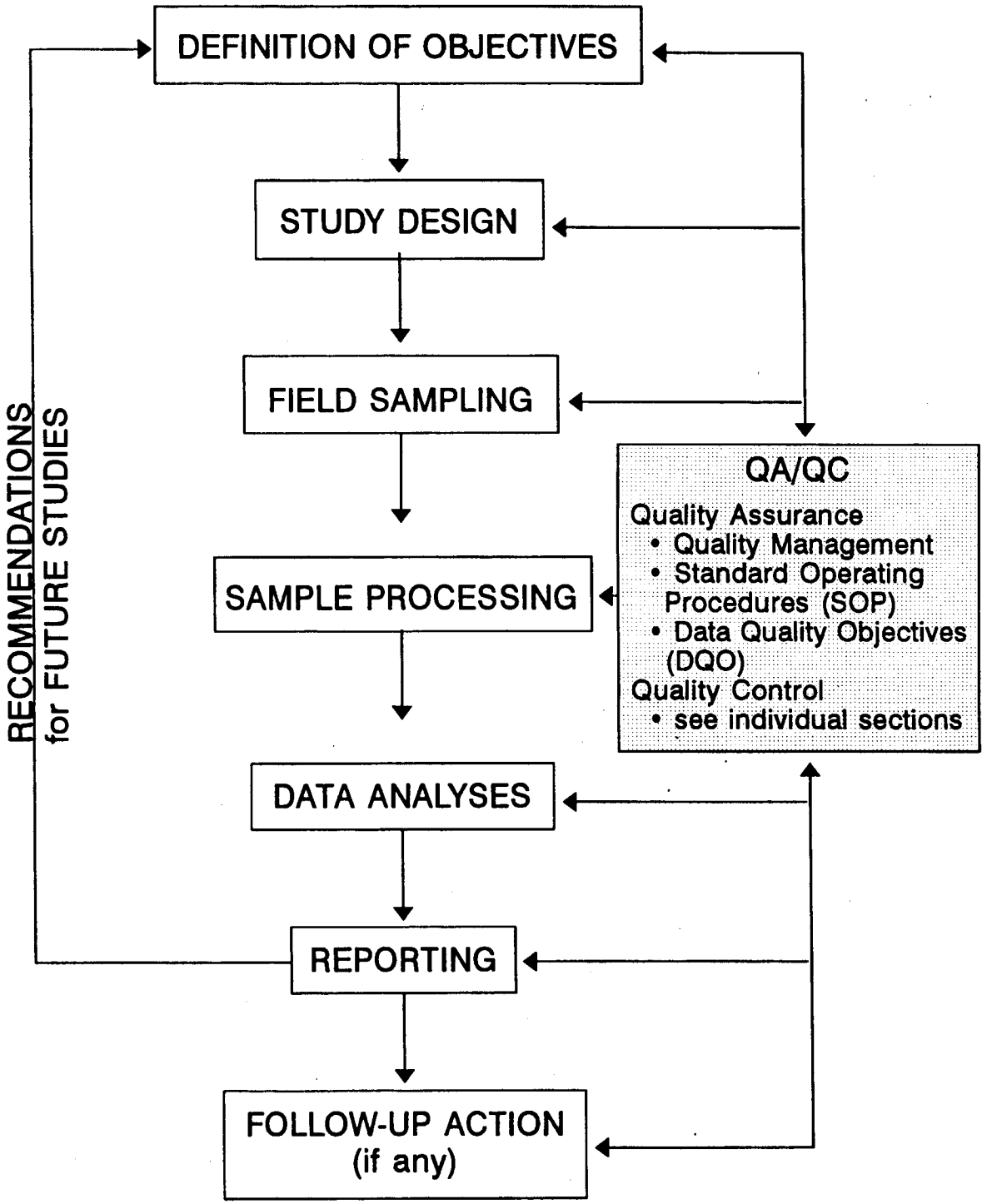

Figure 2. The key elements of a benthos monitoring study. Specific components related to Quality Assurance/Quality Control have been highlighted.

2.0 QUALITY ASSURANCE/QUALITY CONTROL (QA/QC)

QA/QC is an important part of any monitoring program, and applies to all aspects of the program (Figure 2). Users of monitoring data must be confident that those data meet the objectives of the study and are of high quality, particularly if the data are used for regulatory or management purposes.

Quality assurance (QA) refers to externally imposed technical and management practices which ensure the generation of quality and defensible data commensurate with the intended use of the data. Specifically, QA is a set of operating principles that, if strictly followed, will produce data of known and defensible quality (APHA, 1989). For most benthos monitoring programs, the end product must be a set of data that, following analysis and interpretation, will be used to assess whether changes in benthic communities indicate an impact on the receiving environment.

Quality control (QC) is a specific aspect of quality assurance and refers to the internal techniques used to measure and assess data quality and the remedial actions to be taken when the data quality objectives are not realized (APHA, 1989; Environment Canada, 1991). The assurance of data quality is only possible when Data Quality Objectives (DQO) have been defined and followed (see Section 2.1.2).

As an initial step in the development of recommended QA/QC procedures for benthos monitoring programs, the QA/QC procedures in the draft guidance manual for environmental effects monitoring (EEM) for pulp and paper mills (Environment Canada, 1991) were reviewed during the Benthos Monitoring Workshop. In general, the information presented in the EEM document was considered adequate for benthos monitoring studies. However, the workshop did provide modifications, deletions and additions to the checklist of requirements. This section briefly describes general QA/QC procedures and provides a revised checklist of QA/QC requirements developed during the workshop. The QA/QC requirements in this and other sections should be regarded as the minimum necessary to ensure that data are of high quality.

2.1 Quality Assurance

2.1.1 Quality Management

Quality management ensures that QA/QC is stressed at each stage of the monitoring study. It requires a QA/QC Officer to organize procedures and to assign specific individuals to QA/QC functions. Channels of communication must be established and corrective action must be initiated by the QA/QC officer when QA/QC guidelines are not met.

The quality management organizational structure should be documented in a Quality Management Plan (QMP). The QMP should also include a statement of QA policy signed by a Director of budgetary authority. This statement ensures the allocation of resources to quality assurance, and support to all staff in placing quality first.

Standard Operating Procedures (SOPs) and Data Quality Objectives (DQO) should be included in the Quality Management Plan. A brief summary of each is provided below. For detailed information regarding Quality Management Plans, see USEPA (1980) and USEPA (1986).

2.1.2 Standard Operating Procedures (SOP)

Standard operating procedures outline detailed methods for sample collection or analysis. The SOP document contains technical information on specific methods along with the appropriate citation of references, such as published protocol and method manuals. For benthos monitoring studies, SOP will usually contain information regarding sample collection, field procedures, laboratory procedures (sample processing) and analysis. Workshop participants recommended that SOP be included as part of the Quality Management Plan and that SOP be adhered to throughout the study period. Any changes to SOP, or any SOP not commonly used in other studies, should be thoroughly documented and justified. It was obvious from the workshop that there is considerable legitimate debate about the most appropriate procedures for many aspects of benthos monitoring studies. Therefore, SOP should be constantly and critically evaluated, or there is a danger that standardization of procedures will lead to standardization of mediocrity.

2.1.3 Data Quality Objectives (DQO)

Determination of data quality is accomplished through the development of DQO. The DQO are statements defining the level of uncertainty or quality of the data required to meet the study objectives (Klemm et al., 1990). DQO are target values for data quality, not necessarily criteria for the acceptance/rejection of data. Issues to consider when developing the DQOs should include:

- definition of objectives and hypotheses
- study design (including power analysis and the level of uncertainty)
- *a priori* statistical design and decision criteria
- alternatives and contingency plans
- action plans based on study results

2.2 Quality Control

Recommended QA/QC requirements for specific components of a benthos monitoring study are summarized in Table 1. This table is a modification of Table 2.1 included in the EEM draft document (Environment Canada, 1991), and was

modified during the benthos monitoring workshop. Specific QA/QC procedures for the components of monitoring studies shown in Figure 2 are given in Sections 3.0 to 8.0.

2.3 References Cited

APHA (American Public Health Association). 1989. Standard Methods for the Examination of Water and Wastewater. 17th edition. American Public Health Association, American Water Works Association and Water Pollution Control Federation.

Environment Canada. 1991. Technical guidance manual for aquatic environmental effects monitoring at pulp and paper mills, Volume 2: Methodology (draft version). Prepared by Beak Consultants, Toronto, Ont. 199 p.

Klemm, D.J., P.A. Lewis, F. Fulk and J. M. Lazorchak. 1990. Macroinvertebrate field and laboratory methods for evaluating the biological integrity of surface waters. EPA 600/4-90/030. U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Cincinnati, OH.

U.S. EPA (Environmental Protection Agency). 1980. Guidelines and specifications for preparing quality assurance project plans. QAMS-005/80. U.S. Environmental Protection Agency, Office of Monitoring and Quality Assurance, Office of Research and Development, Washington, D.C.

U.S. EPA (Environmental Protection Agency). 1986. Development of data quality objectives. Description of stages I and II. Prepared by the Quality Assurance Management Staff, U.S. Environmental Protection Agency, Office of Research and Development, Washington. D.C.

Table 1. QA/QC checklist for benthos monitoring studies (modified from Environment Canada, 1991).

QUALITY MANAGEMENT

___ Quality Management Plan prepared.
___ Standard Operating Procedures available (sampling, processing and QA methods, revisions).
___ Responsibilities, authorities, qualifications defined for each position.
___ Quality Assurance Officer with adequate authority for corrective action.
___ Data Quality Objectives for study (hypotheses, study designs, sensitivity, accuracy/precision, total uncertainty).

QUALITY ASSURANCE FUNCTIONS PERFORMED

___ Review of QC sample results (permanent record, replicates, exchanges, verifications).
___ Performance assessment review (blind samples).
___ Knowledge testing and training sessions (field and laboratory).
___ Review of test sample data (transcription and logic checks performed).
___ Report approval mechanisms (signatures).
___ Quality Assurance reports (corrective actions indicated).
___ Interlaboratory and accreditation studies (record of participation).

FIELD OPERATIONS

___ Preliminary investigations (background information, reconnaissance survey).
___ Sampling design (based on objectives, adequate replication).
___ Consistency in sampling methods throughout study period (and with previous studies)
___ Instrument calibration and maintenance (records, methods available).
___ Staff training and evaluation/experienced personnel in the field.
___ Samples must be collected correctly (sediment penetration, samples intact, undisturbed)
___ Sampling equipment (appropriate, consistent, cleaning and maintenance records).
___ Collection, preservation, shipping, storage (methods, adequate labels, custody records).
___ Field notes maintained (accurate site locations, habitat descriptors, substrate type, flow, etc.)

SAMPLE PROCESSING

___ Sorting and subsampling (methods, adequate records, error estimates).
___ Sorting verification (confirm few organisms overlooked).
___ Taxonomy and enumeration (appropriate keys, records, reference collection).
___ Documented sorters, and identifiers' names and qualifications
___ 95% recovery of organisms (sorting QA/QC).
___ Taxonomic verification (by recognized experts).
___ Archives (for samples, voucher specimens).

Table 1. (continued)

REPORTING

___ Data handling and reporting (entry checks, missing value codes, methods, QC data).
___ Detailed report of all methods (collection, processing, analyses).
___ Detailed reporting of all pertinent information (including field notes, map, accurate site locations).
___ Changes in protocol, study design or other components of the study.
___ All QA/QC documentation included in the report (as an appendix).
___ All raw data (biological, chemical, physical) in appendices.
___ Review of completed study report

Notes/Comments:

_____

DEFINITION OF OBJECTIVES

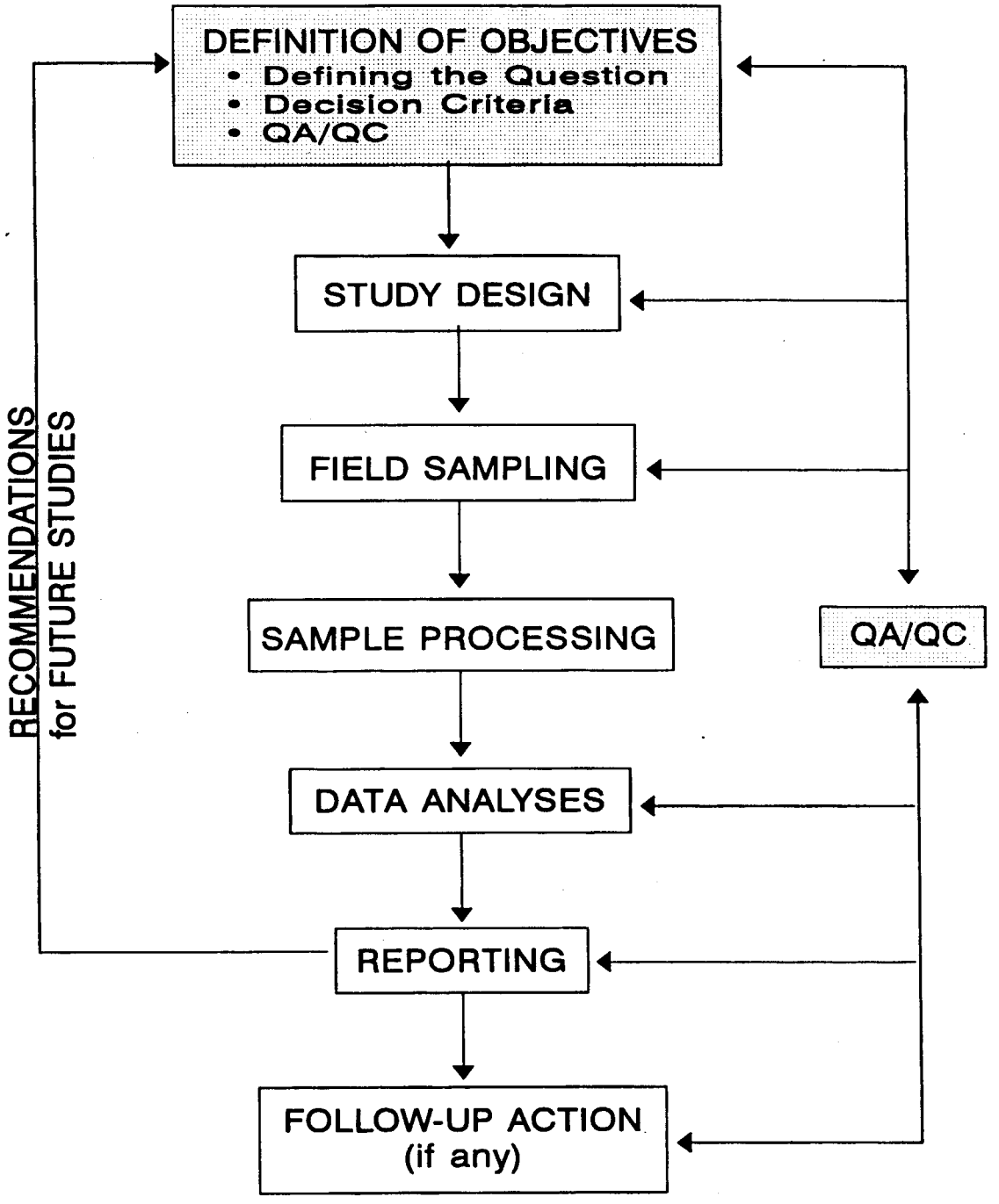

Figure 3. The key elements of a benthos monitoring study. Specific components related to defining study objectives have been highlighted.

3.0 DEFINING THE QUESTION

Before a benthos monitoring study can be initiated, the objective(s) of the study must be defined as clearly and concisely as possible and phrased as a testable question or hypothesis. Although benthos monitoring studies can have diverse objectives, most, if not all, can be defined as spatial, temporal or spatio-temporal questions or hypotheses. Spatial studies investigate changes in benthos over spatial areas or distances (e.g., upstream vs. downstream of an outfall). Temporal studies investigate changes in benthos over time (e.g., operational vs. baseline data). For long-term studies, the optimal questions would incorporate both spatial and temporal comparisons. Examples of spatial and temporal hypotheses, typically tested in biomonitoring studies investigating possible effects of point-source pollution, are provided below. These examples have been taken from the draft technical guidance manual for the pulp and paper Environmental Effect Monitoring program (Environment Canada, 1991) and refers to comparisons of a single variable (e.g., abundance or richness):

Spatial:

- Does the exposure area response mean differ from the reference mean?
- Do the near-field station means differ from far-field station means?
- Does the response mean for any station violate regulatory guidelines?

Temporal:

- Has the response mean changed since the last study?
- Has the response mean changed from pre-operational or baseline conditions?
- Has the magnitude of difference between the exposure and reference means changed?

The objectives lead to the formulation of specific testable hypotheses or questions to be answered by the monitoring program. An hypothesis represents a further refinement (increased precision) of an objective statement by "including within it information about the criterion [e.g., a measure of biological impact] and predictor variables [e.g., measure of impact intensity]" (Green, 1979). The hypothesis should represent the simplest possible answer to the objective stated so that it is testable and falsifiable (i.e., the null hypothesis H_o; see Green (1979) for more information on hypothesis formulation).

3.1 Decision Criteria

Many benthos monitoring studies are conducted in response to specific regulatory requirements. Consequently, the regulator may have outlined the general objectives of the study and, perhaps, the specific hypotheses to be tested. The objectives and hypotheses obviously must be reviewed by regulators, investigators and other involved parties (e.g., industry) before the study has been initiated. In addition, the regulator should provide the criteria that will be used to decide whether a change in the exposed benthos community is considered biologically significant and represents an impact on the receiving environment (i.e., what magnitude of change in exposed benthic communities is considered an impact and warrants mitigative action). These criteria may be developed on a site-by-site basis and must be reviewed and understood before the objectives and specific hypotheses of the study have been finalized. The decision criteria are important in defining the level of resolution of the study required by the regulator so that an informed decision can be made regarding the confidence of the data and the detection of significant impacts. The investigator should indicate whether the criteria are realistic and the required resolution achievable. It should be emphasized, however, that a statistically significant change in benthic communities does not necessarily indicate a meaningful or important change (i.e., have ecological or human consequences) (NRC, 1990). This issue is also discussed in Section 7.3 (Data Presentation and Interpretation).

3.2 Quality Assurance/Quality Control

The objectives and specific hypotheses should go through a final review process involving the regulator, industry and knowledgable experts (e.g., biometrician). The review process must ensure that the objectives of the study are clear to all parties and that the hypotheses meet the objectives and are testable (i.e., avoid open-ended questions which are often characterized as "fishing expeditions"). Decision criteria used to interpret the study results and initiate mitigative action (e.g., treatment system, legal action, etc.) must be documented.

3.3 References Cited

Environment Canada. 1991. Technical guidance manual for aquatic environmental effects monitoring at pulp and paper mills. Volume 1: Overview and study design (draft report). Prepared by Beak Consultants. 97 pp.

Green, R.H. 1979. Sampling Design and Statistical Methods for Environmental Biologists. John Wiley and Sons, Toronto. 257 pp.

NRC (National Research Council). 1990. Managing Troubled Waters: The Role of Marine Environmental Monitoring. National Academy Press, Washington, D.C. 125 pp.

STUDY DESIGN

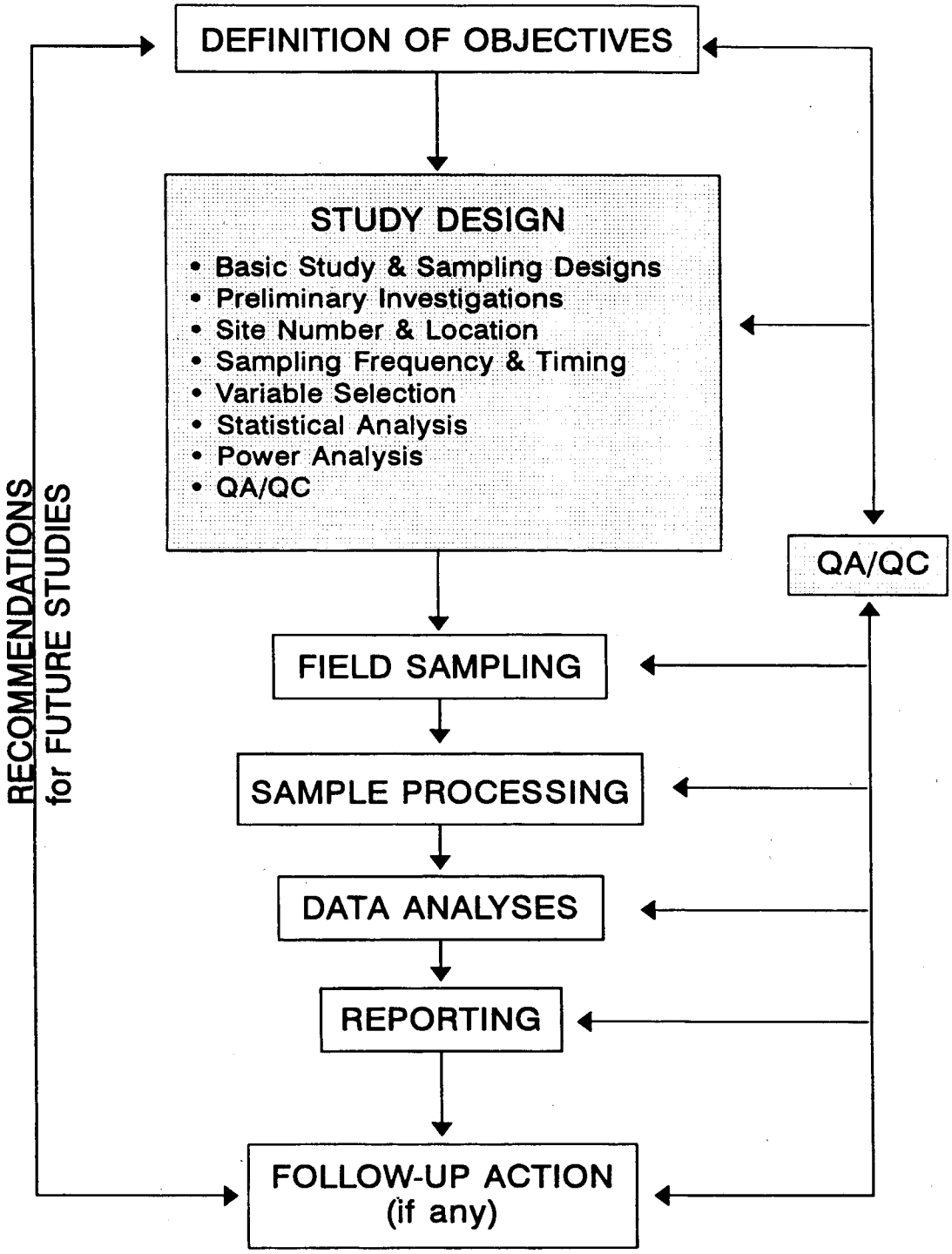

Figure 4. The key elements of a benthos monitoring study. Specific components of a study design have been highlighted.

4.0 STUDY DESIGN

This section reviews the major elements of designing a study to assess benthic communities in freshwater habitats. The focus is on descriptive surveys or natural experiments, rather than manipulative experiments. In descriptive surveys or natural experiments, the investigator cannot directly manipulate the factors of interest. For example, if the objective is to assess the effects of an industrial discharge on benthic communities, the investigator cannot directly manipulate or control the presence or absence of the discharge at selected sample sites or times. Instead, samples would be taken upstream and downstream of the discharge, or before or after the discharging industry becomes operational (if possible). In a manipulative experiment, the investigator can manipulate the factor of interest while holding all other factors constant. As a result, stronger inferences about cause and effect can be drawn from manipulative experiments. Some authors (e.g., Hairston, 1989) have argued that natural experiments should not even be called experiments, but the term is used here to indicate that many of the approaches used in manipulative experiments can be adapted to the design and analysis of natural experiments. Investigators conducting monitoring studies should be aware of the limitations of observational surveys and natural experiments, but should also be aware that inferences from their studies can be strengthened by following the guidelines in this section and in Section 7.0 (Data Analyses).

Depending on the objective of the study, manipulative experiments may be required. Mesocosms were advocated during the workshop as an experimental approach which could be useful in impact assessment. It was suggested (with mixed reviews) that mesocosm experiments would be effective in determining cause-and-effect relationships between possible stressors (e.g., contaminants) and the response of benthos. Although this document does not focus on design for manipulative experiments in constructed environments (e.g., mesocosms), information on mesocosms presented at the workshop is summarized in Appendix A.

The following sections present the key elements of study design discussed during the workshop.

4.1 Basic Study and Sampling Designs

The general study design must reflect the objectives and hypotheses of the monitoring study. Consequently, the design of a study will typically be one of three types (Green, 1979; Underwood, 1991):

1) Spatial or Control-Impact (CI) Design - the most common design which involves a comparison of benthic communities between potentially impacted sites and reference or control sites. Differences between the reference/control and impact sites are considered indicative of effects.

2) Temporal Design - comparisons of benthic communities over time. Frequently, benthic communities are compared before and after a discharging industry becomes operational, or before and after some change (e.g., in effluent treatment or management/regulatory practice), and the design is referred to as Before-After (BA). Less frequently, benthic communities are monitored over a long time period as an indicator of overall environmental health (i.e., long-term trend monitoring).

3) Site-by-time or Before-After-Control-Impact (BACI) Design - a combination of the spatial (CI) and temporal (BA) designs. The *difference* between control and impact sites is compared before and after some event (e.g., operation of a discharging industry; implementation of treatment improvements or discharge reduction).

There are also other variations on these basic designs (e.g., repeated measures design, nested design, block designs, etc.), which may be more applicable to specific study objectives or study areas. The design of the study (and hypothesis) is also important in determining the appropriate statistical test to be used to analyze the data. More specific information regarding data analyses is provided in Sections 4.8 and 7.2.

The study designs provided above have been criticized as examples of pseudoreplication (e.g., Hurlbert, 1984). In a manipulative experiment, treatments are assigned randomly to experimental units or replicates (i.e., each replicate has an equal chance of receiving any particular treatment), but this is impossible in most monitoring studies. For example, a treatment such as the presence of effluent contaminants cannot be assigned to replicate samples taken upstream of the discharge or at times prior to the operation of the discharge. As a result, a difference between sites or times is not necessarily indicative of a contaminant effect, as there are many other factors which might cause such a difference. The evidence for contaminant effects can be strengthened by eliminating the potential effects of other factors, and methods for doing that are presented in this section and in Section 7.0 (Data Analyses).

The simplest spatial or CI design consists of a comparison of benthic communities sampled at a reference or control site with communities sampled at an impact or exposed site (e.g., influenced by effluent discharge). Although simple, this basic design may be adequate for the majority of benthos monitoring studies investigating effects of point-source pollution, if modified regarding the number of reference and exposed sites sampled, the location of sites, and other considerations as discussed below. The technical guidance manual of the EEM program (Environment Canada, 1991) provides a generalized sampling design that incorporates the simplified design and further subdivides the exposed sampling area. This document recommends that sampling sites be located in at least three spatial areas or regions:

1) Reference area: an area that is not exposed to the potential source of pollution but exhibits similar natural characteristics to the exposed monitoring sites.

2) Near-field exposure area: an area of high exposure to the source of potential impact but beyond the immediate region of discharge.

3) Far-field exposure area: an area of less exposure than the near-field area but still within the zone of influence.

If possible, a fourth (recovery) area even further from the source should be sampled. These three or four areas will generally define a gradient of contamination. The actual location of these sampling areas, relative to the source of contaminant, will be based on site-specific dilution factors and predicted effects. Note that if the sites represent a gradient of contaminant concentration, and if effects are correlated with this gradient, then the evidence for contaminant effects is increased.

If a BA or BACI design is used, samples must obviously be taken before and after some event, and the occurrence of that event will determine the sample times. As with CI designs, the study will be improved if samples are taken at several different times both before and after the event (Underwood, 1991). Depending on the type of statistical analyses used, sample times can be at regular intervals (e.g., same time every year; monthly within a year) or randomly selected.

4.2 Preliminary Investigations

Preliminary investigations refine site selection, select sample sizes and sampling frequencies, and evaluate the parameter list and sampling methods.

4.2.1 Review of Background Information

Prior to designing the study, it is first necessary to review relevant historical data and literature. Adequate historical information may fulfil pre-design requirements, reduce the level of effort and overall costs of the program, and focus and refine the study design and sampling methods. Previous studies of similar problems or sites may provide appropriate designs and methods. Common sources of information include maps, previous reports on the study area, and physical, chemical and biological data records.

Specific information useful to the study could include (but is not restricted to):

- maps and descriptions of the study area
- information regarding potential pollution sources (e.g., industrial operations data, effluent chemistry, etc.)
- resource and habitat inventories
- receiving water chemistry
- effluent plume delineation
- sampling variability
- methods used in previous studies

If historical information is limited or not available, a reconnaissance survey should be conducted prior to the benthos monitoring study.

4.2.2 Reconnaissance Survey

A reconnaissance survey was strongly recommended during several workshop discussions. Such a survey was considered a cost-effective means to familiarize the researchers with the study area and to collect necessary data not available in historical literature. The following recommendations apply:

- review maps and air photos of the specific study area to plan the reconnaissance survey
- review historical data (physical, chemical, and biological) to determine how much new information needs to be gathered during the reconnaissance
- conduct a plume delineation study if one has not already been done
- conduct a habitat inventory of the study area
- map the location of riffles, deposition zones, potential pollution sources (point and/or non-point), relevant land use, access points and tributaries
- document the substrate type, flow characteristics or any other factors that may influence the natural benthos community or the ability to sample those communities
- conduct a preliminary water quality assessment (e.g., pH, dissolved oxygen, temperature, conductivity, and other on-site measures)
- conduct a preliminary sampling of benthic communities to:
    - select reference and exposed sites
    - provide an estimate of background variability to be used for power analysis calculations (see Section 4.7)
    - test the sampling approach
    - identify resident organisms, and verify these identifications
- test equipment and methods.

4.3 Site Number and Location

In a CI or BACI design, several sites should be sampled within each of the areas defined in Section 4.1, to encompass the spatial variation of the aquatic system and span the gradient of contamination. The actual number of sites sampled will depend on the specific study area (e.g., tributaries, additional discharges, habitat) and, to some extent, the resources available. The EEM program (Environment Canada, 1991) has recommended at least three stations within each sampling area (provided historical or background data are available). The workshop recommended that a minimum of two reference sites be sampled.

Sites should be as similar as possible in all characteristics except the factor of interest (e.g., distance from discharge source). Variation in substrate type, gradient and current velocity (streams), depth, and riparian habitat or other physical or chemical attributes should be minimal or removed during data analysis (e.g., by using blocks or covariates; see Section 7.0). Sites should not be located where the benthos community may be influenced by atypical conditions such as bridges, channelization, dredging or culverts, unless these are the potential impacts of interest.

The type of habitat (e.g., depositional *versus* erosional habitat) that should be sampled has frequently been the subject of debate. Traditionally, in streams, the majority of benthic invertebrate monitoring has focused on riffle communities because of ease of sampling, increased sampling precision, higher diversity in riffles, the presence of pollution-sensitive taxa, the presence of sport fish, and the abundance of supporting literature describing the expected characteristics of riffle communities. However, contaminants may be transported past a riffle habitat to accumulate in the sediments of depositional areas such as pools. Therefore, invertebrates in pool habitats may experience more direct and consistent exposure to contaminants than riffle communities. In toxic substance studies, it is important to consider the nature and duration of the probable exposure for all benthic invertebrate communities sampled. The EEM document (Environment Canada, 1991) suggests that each dominant, well represented habitat type should be sampled. Separate reference, near-field, far-field and recovery sites would then be sampled for each habitat type. This strategy circumvents many of the habitat-specific problems mentioned but also increases costs. The characteristics of the study area, type of potential pollution source and the objectives of the study will dictate the approach taken.

Reference sites (i.e., control stations) should be located in areas not exposed to sources of contamination. In rivers, this is typically upstream of the outfall ensuring that there are no additional sources of contaminants between the reference site(s) and the outfall. If upstream sites are not available, reference sites can be located in unexposed downstream locations (similar to recovery sites) or in adjacent reference streams of similar physical and chemical characteristics. In lakes, reference sites may be located upshore or downshore of the outfall location (depending on prevailing currents), in similar habitats but on the opposite or distant shores, or in other reference lakes of similar limnological characteristics (Environment Canada, 1991). In both rivers and lakes, plume delineation studies examining the dispersion and dilution of dissolved and particulate material are invaluable for indicating the most appropriate reference and potentially impacted sites.

4.4 Sampling Frequency and Timing

Most of the following information has been taken from a recent protocol developed by Alberta Environment (1990) and the draft EEM document (Environment Canada, 1991), although some of these ideas were discussed during the workshop.

The following factors should be considered when deciding on the frequency and timing of benthos sampling:

- specific objectives of the study - spatial or temporal design, monitoring seasonal or annual changes in benthic communities

- the benthic communities - times of highest diversity (e.g., fall); times when there are fewer immature individuals that are difficult to identify (e.g., spring); times when benthos are most likely to show the strongest response to discharges (high temperature and low flow or dilution); the life history of the organisms

- habitat and climatic condition - onset of winter conditions and considerable ice-cover; hydrological regime

- type of pollution source discharge - continuous or discontinuous discharge; changes in the quantity and quality of effluent; accidental discharges.

4.5 Variable Selection

It was recommended that benthos monitoring studies select variables that described the following equally important components of the study:

1) the benthos community (e.g., taxonomic richness, taxa abundances, biomasses etc.)

2) the proximity to the potential impact (e.g., distance, contaminant concentration(s), etc.)

3) the nature of the non-impacted habitat (e.g., physical and chemical properties).

The selection of specific variables to describe or quantify benthic communities will reflect the study objectives and hypotheses. However, many previous studies focused on abundance and presence/absence data by default. The question of "how" to appropriately quantify benthic communities was brought up during the workshop but was not completely resolved. The merits of various variables and metrics are discussed by Klemm et al. (1990) and Resh and Jackson (1993).

4.6 Choice of Statistical Analysis

The choice of statistical analysis must reflect the study objective, hypothesis, experimental design and sampling design. Selecting the appropriate analysis should be part of the study design development rather than an *ad hoc* decision occurring after the samples have been collected. According to Green (1979), an efficient statistical analysis should be:

- conservative - have a low probability (") of making a Type I error (i.e., concluding that there is an effect when there is none)

- powerful - have a low probability ($) of making a Type II error (i.e., concluding that there is no difference when one exists)

- robust - the stipulated error levels will not be seriously affected by the different kinds of data encountered in environmental studies.

For detailed information regarding data analysis and appropriate statistical tests, see Section 7.2.

4.7 Power Analysis

Statistical power concepts are used to develop and evaluate study designs. After determining the location and number of sample sites or times, and the potential measurement variables, the investigator must determine the number of samples (replicates) collected at each site or time. Typically, 3-5 samples per site or time are used as a rough compromise between precision and costs, often without considering information on sampling variance and other factors. Power analysis is recommended as an objective method to determine the number of samples required. Power refers to the probability that an effect (i.e., difference or change) of a given size will be detected, if the effect actually exists. The most common application of power analysis is calculation of the required sample sizes to achieve the desired power. The power of different sampling designs can also be compared to select the most appropriate in terms of power per unit cost (e.g., Ferraro et al., 1989). The power of different variables can also be compared, and the most powerful selected for measurement. The benefits of using monitoring designs with high power are increased confidence that effects will be detected if they exist, and more efficient use of financial resources.

Power calculations consider two probability values (" and $). The first, ", is more familiar and is the probability that the observed differences could occur by chance alone. Scientists normally set " at 0.05 or 0.10 to guard against declaring that a difference exists when it does not (Type I Error). However, it is also possible to declare that a significant difference does not exist when, in fact, it does. This is known as the Type II Error, and the probability of making such an error is $. Power is simply 1-$ (or 100-$, if probabilities are expressed as percentages). Statistical power increases with increasing sample size, effect size and ", and decreases with increasing variance among replicates.

There has been increasing concern with reducing $, as well as ", in environmental assessments and monitoring programs (e.g., Alldredge, 1987; Peterman, 1990). The reason for keeping " low is obvious: to make sure that a difference really exists between, for example, benthic communities upstream and downstream of an effluent discharge, before recommending action such as installing expensive treatment systems or taking legal action against polluters. However, it is also important to keep the value of $ low. If it is concluded, for example, that a discharge has no effect on benthic communities, when in fact it does, this is not protecting the communities or receiving environment.

There are many references that can be referred to regarding power analysis and calculating sample sizes and power (e.g., Skalski and McKenzie, 1982; Bernstein and Zalinski, 1983; Alldredge, 1987; Green, 1989; Peterman, 1990), so formulae will not be provided in detail here. Goldstein (1989) provides a summary of computer programs which can be used for power analysis. Investigators conducting benthos monitoring programs should use power analysis prior to sampling to calculate sample sizes and compare sampling designs, and should also report the minimum detectable effect size in the final report.

Much of the discussion at the workshop, and some of the references cited in this section, deal almost exclusively with the simple case of a univariate *t*-comparison between one reference and one impacted site. However, the formulae and principles can be adapted for other types of designs and tests. For example, *n* (sample size per site) for any contrast (see Section 7.2) can be calculated by multiplying *n* for the simple *t*-comparison by $\Sigma 8^2/2$, where the 8 are the contrast coefficients. This conversion was derived by substituting the standard error of a contrast for the standard error of a difference (Snedecor and Cochran, 1980). One problem with using power analysis for contrasts is that power (or required *n*) will differ among contrasts if $\Sigma 8^2$ differs. Investigators should also be able to calculate sample sizes for other methods such as multiple range tests.

Multivariate analyses present more difficult problems for power calculations, primarily because effect sizes cannot be specified. The power of multivariate analyses relative to univariate analyses depends on the correlations among the variables, differences in response among variables, and whether the vector of response differs from the dominant vector of natural variation. Green (1989) provides a good discussion of the problem, provides methods for placing some bounds on the problem, and also mentions computer programs available for multivariate power calculations and simulations.

4.8 Quality Assurance/Quality Control (QA/QC)

Before the monitoring study is implemented, the study design must be reviewed and approved by the researcher, regulator and a knowledgable expert (e.g., biometrician). The review should evaluate:

- how well the design meets the study objectives and test the hypotheses
- the technical elements of the design
- the statistical elements of the design (including statistical power)

- the economic costs (both the cost of sampling and the cost of the outcome)

- the overall feasibility of the proposed design (technical, logistic, statistical, economic, etc.), considering appropriate modifications or possible alternative designs (if needed) that would reflect the study objectives.

4.9 References Cited

Alberta Environment. 1990. Selected methods for the monitoring of benthic invertebrates in Alberta rivers. Water Quality Branch, Pollution Control Division. 41 pp.

Alldredge, J.R. 1987. Sample size for monitoring of toxic chemical sites. Envir. Monitoring Assess. 9:143-154.

Bernstein, B.B. and J. Zalinski. 1983. An optimum sampling design and power tests for environmental biologists. J. Envir. Mgmt. 16:35-43

Environment Canada. 1991. Technical guidance manual for aquatic environmental effects monitoring at pulp and paper mills. Volume 1: Overview and study design (draft report). Prepared by Beak Consultants, Toronto, Ont. 97 pp.

Ferraro, S.P., F.A. Cole, W.A. DeBen and R.C. Swartz. 1989. Power-cost efficiency of eight macrobenthic sampling schemes in Puget Sound, Washington, USA. Can. J. Fish. Aquat. Sci. 46: 2157-2165.

Goldstein, R. 1989. Power and sample size via MS/PC-DOS computers. Amer. Statist. 43: 253-260.

Green, R.H. 1979. Sampling Design and Statistical Methods for Environmental Biologists. John Wiley and Sons, Toronto, Ont. 257 pp.

Green, R.H. 1989. Power analysis and practical strategies for environmental monitoring. Environ. Res. 50:195-205.

Hairston, N.G. Sr. 1989. Ecological Experiments: Purpose, Design, and Execution. Cambridge University Press, Cambridge, UK. 370 pp.

Hurlbert, S.J. 1984. Pseudoreplication and the design of ecological field experiments. Ecolog. Monogr. 54: 187-211.

Klemm, D.J., P.A. Lewis, F. Fulk and J.M. Lazorchak. 1990. Macroinvertebrate field and laboratory methods for evaluating the biological integrity of surface waters. EPA 600/4-90/030. U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Cincinnati, OH.

Peterman, R.M. 1990. Statistical power analysis can improve fisheries research and management. Can. J. Fish. Aquat. Sci. 47:2-15.

Resh, V.H. and J.K. Jackson. 1993. Rapid assessment approaches to biomonitoring using benthic macroinvertebrates. pp. 195-233. In: Freshwater Biomonitoring and Benthic Macroinvertebrates. D.M. Rosenberg and V.H. Resh (eds). Chapman and Hall, New York, NY.

Skalski, J.R. and D.H. McKenzie. 1982. A design for aquatic monitoring programs. J. Envir. Mgmt. 14:237-251.

Snedecor, G.W. and W.G. Cochran. 1980. Statistical Methods. 7th Ed. Iowa State University Press, Ames, Iowa. 507 pp.

Underwood, A.J. 1991. Beyond BACI: experimental designs for detecting human environmental impacts on temporal variations in natural populations. Aust. J. Mar. Freshwat. Res. 42: 569-587.

FIELD SAMPLING

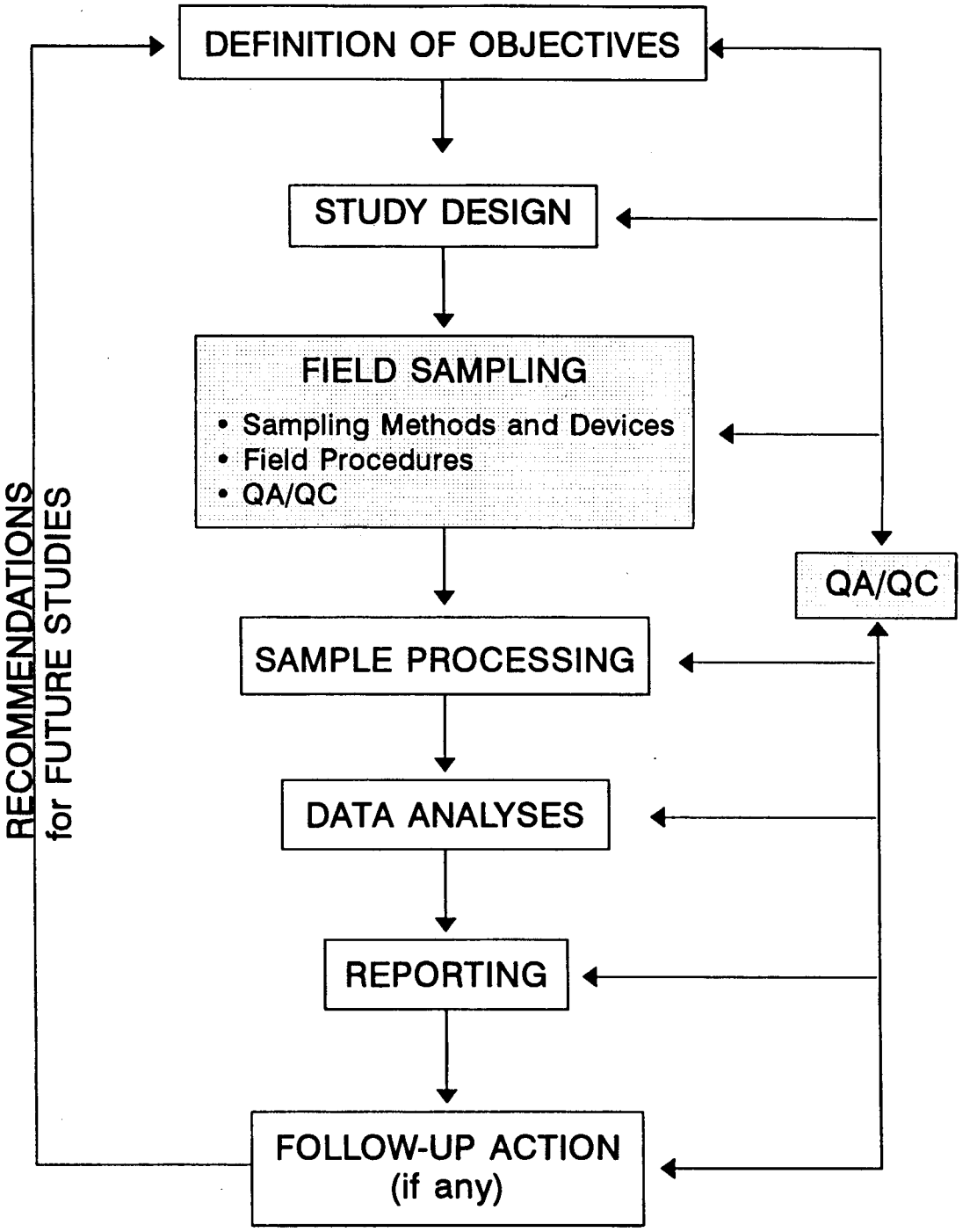

Figure 5. The key elements of a benthos monitoring study. Specific components of field sampling have been highlighted.

5.0 FIELD SAMPLING

This section outlines the basic elements of field sampling methods and devices, field procedures, and related Quality Assurance/Quality Control (QA/QC) considerations. Prior to selecting a sampling method, it is critical to review past studies performed in the vicinity of the study area or in similar areas. This information will allow more informed decisions about what sampling devices are most efficient, and selection of optimal field procedures.

5.1 Sampling Methods and Devices

Benthic invertebrates can be collected from either natural or artificial (introduced) substrates with each type offering advantages, depending on site-specific conditions. Natural substrates should be sampled wherever possible. Artificial substrates should be used when natural substrate cannot physically be sampled or when the substrate is so variable that the effect needs to be removed from the sampling design.

5.1.1 Mesh Size

In almost every benthos monitoring study, sampling devices with screens or nets, or sieves, are used. As a result, the smallest invertebrates will not be retained. The mesh size chosen will determine the minimum size of organisms sampled, and the degree to which abundance is underestimated. Smaller mesh sizes will retain more organisms and provide a better estimate of abundance, but will also increase the costs of sample processing. A mesh size of 180-250 Fm is recommended for benthos monitoring studies, unless there is a particular interest in focusing on smaller (e.g., Oligochaeta) or larger (e.g., Plecoptera or Ephemeroptera) organisms. If samples taken in one study are to be compared to those taken in an earlier study, the mesh size should be the same in both studies.

5.1.2 Natural Substrates

For the majority of monitoring studies, natural substrates are the preferred habitats to sample. The major advantage of samples from natural substrate is that they reflect the indigenous benthic invertebrate community structure. Potential disadvantages include higher sampling variance resulting from natural substrate heterogeneity, which may in turn increase costs through requirements for large sample sizes (Klemm et al., 1990). Natural substrates in freshwater can be sampled using a wide variety of devices, including grab, stream-net, core, and air-lift (suction) samplers. Some authors have sampled single stones, rather than fixed areas (e.g., Doeg and Lake, 1981; Wrona et al., 1986; Scrimgeour et al., 1993). The consensus among workshop participants was that large, fast rivers with relatively coarse substrates were probably the most difficult freshwater habitats to sample, but there was little agreement on the most effective methods. Grab and stream-net samplers are recommended for most freshwater monitoring studies and habitats. The characteristics of various samplers are summarized in Table 2. If used properly, these samplers provide quantitative estimates of abundances per

unit area, but it is important to realize that the abundance of small organisms will be seriously underestimated because of loss through net or sieve mesh.

Grab samplers collect a sample by penetrating the substrate and obtaining a discrete quantity of bottom sediment. All grab samplers have a jaw mechanism that i) closes upon impact with the sediments, or ii) tripped with a messenger. Rocks and gravel can get caught in the jaws, preventing them from closing, and leading to loss of the sample. Therefore, grabs should only be used for sampling finer substrates. Because the grabs can be operated from a boat, and triggered using the messenger, they are suitable for deeper waters where stream-net samplers cannot be used. The tops of most samplers have hinged lids, which are open when the sampler descends but close when the sampler is hauled up. However, active invertebrates may easily escape unless the lids form a perfect seal. Therefore, the tops of grab samplers should be covered with 180-250 Fm mesh. The Ponar (standard and petite), van Veen and Ekman samplers are recommended for assessing benthic invertebrate communities (Table 2).

Stream-net samplers are fitted with a fine-mesh net and collect benthic invertebrates from flowing water passing through the sampler. These samplers are typically used in shallow waters (< 0.5 meters) with coarser substrates typical of riffle habitats. The recommended stream-net samplers include the Hess, Box and Surber samplers, and other samplers of similar design (Table 2).

5.1.3 Artificial Substrates

An artificial substrate is defined as any introduced device used to standardize substrate features of the aquatic environment into which it is placed. Artificial substrates can be used to monitor changes in invertebrate communities over time and space, but do not necessarily reflect the benthic invertebrate community that resides in and on the natural substrate. The communities colonizing artificial substrates will be biased towards mobile or drifting organisms. Estimates of abundance obtained from artificial substrates should be expressed as numbers per sampler, because they are not estimates of abundance on adjacent natural substrates.

The advantages and disadvantages of artificial substrates are given below (summarized from Klemm et al., 1990, and comments from workshop participants):

Advantages:

- allow collection of data from locations that cannot be sampled effectively by other means

Table 2. Summary of recommended grab and stream-net samplers for assessing benthic invertebrate communities in freshwater (APHA, 1989; Klemm et al., 1990; ASTM, 1992).

SAMPLER	SIZE	HABITAT	EFFECTIVENESS	ADVANTAGES	LIMITATIONS
GRAB SAMPLERS					
Standard Ponar (screened)	0.052 m^2	Freshwater lakes, rivers, estuaries, and reservoirs with hard and soft sediments such as clay, hard pan, sand, gravel and muck; somewhat less efficient in softer sediments.	Not entirely adequate for deep burrowing organisms in soft sediments; very efficient for hard sediments; collects both qualitative and quantitative samples.	Better penetration that other grabs; side plates and screens reduce washout, shock waves and substrate disturbance; best quantitative grab sampler for freshwater use.	A very heavy grab that requires use of a boat with winch and cable; stones, pebbles and other debris can hold jaws open causing loss of sample.
Petite Ponar (screened)	0.023 m^2	Freshwater lakes, rivers, estuaries, and reservoirs with moderately hard sediments such as sand, silt and mud; will not penetrate clay; somewhat less efficient in soft sediments and coarse gravel.	Not entirely adequate for deep burrowing organisms in soft sediments; not useful in clay.	Good penetration for such a small grab; side plates and screens reduce washout, shock waves and substrate disturbance; can be operated by hand without boat or winch.	Jaws can be blocked by stones, sticks and other debris causing loss of part of the sample; not efficient in swiftly flowing water of over one meter per second velocity.
Van Veen (screened)	0.06 or 0.1 m^2	Marine waters and estuaries, adaptable to freshwater areas (e.g., for large rivers, deep lakes). Good for sand, gravel, mud, clay and similar substrates.	Penetrates to a depth of 5 to 7 cm.	Jaws close tight; samples most sediment types; comes in a range of sizes.	A very heavy grab that requires a large boat and power winch; jaws may become blocked by debris as in rocks and sticks; not useful for deep burrowing organisms.
Ekman (screened)	0.023 or 0.052 m^2	Freshwater rivers, lakes and reservoirs where there is little current; soft sediments such as muck and silt.	Efficient only in soft sediments but weights can be added for deeper penetration in fine sand; collects both qualitative and quantitative samples.	Easy to operate by hand without winch, can be pushed into substrate in shallow water; hinged doors at top reduce washout, shock waves and disturbance of the substrate; comes in a range of sizes.	Light weight; will not penetrate hard substrates; jaws may close incompletely due to blocking of jaws or failure of closing mechanism; inefficient in deep water or moderate current.

Table 2. (continued)

SAMPLER	SIZE	HABITAT	EFFECTIVENESS	ADVANTAGES	LIMITATIONS
STREAM-NET SAMPLERS[a]					
Hess	0.1 m^2	Shallow, flowing streams, less than 32 cm in depth with good current; rubble substrate, mud, sand and gravel.	Relatively quantitative when used by experienced biologist; performance depends on current, substrate and mesh size.	Completely encloses area samples; prevents escape of organisms; stable platform; samples a unit area; can be used in weed beds.	Difficult to set in some substrate types, that is, large rubble; cannot be used efficiently in still, slow moving waters.
Box	0.1 m^2	Same as above	Same as above	Same as above	Same as above
Surber	0.1 m^2	Same as above	Same as above	Encloses area sampled; easily transported or constructed; samples a unit area.	Same as above.

[a] Stream-net samplers may vary in size depending upon manufacturer. For example, the Surber sampler was historically constructed in the United States as a 1 ft^2 sampler (0.09 m^2); therefore always confirm the dimensions of any stream-net sampler used.

- permit standardized sampling
- may reduce variability compared with other types of sampling
- reduce time for sample processing because there is usually less detritus than in natural substrate samples
- permit greater flexibility in sampling programs.

Disadvantages:

- colonization dynamics not well documented
- sample may be non-representative of local conditions if invertebrates colonizing the samplers originate well upstream of the site
- artificial substrates require long exposure time (6-8 weeks)
- potential loss of fauna on retrieval of samples
- two trips are required; one to place the samplers in the stream or lake, and a second to remove them
- artificial substrates are often lost or vandalized.

There are two main types of artificial substrates commonly used: the multiplate (Hester-Dendy) sampler and the basket sampler (APHA, 1989; ASTM, 1992). Multiple-plate samplers consist of standardized, reproducible surfaces (normally tempered hardboard or ceramic material) for colonization by aquatic organisms. They have a uniform shape and a known surface area. They are used in most aquatic habitats with the exception of wetlands. Multiple-plate samplers are selective for certain groups of invertebrates (e.g., filter-feeders).

While there is no standard basket sampler, the most commonly used basket is a cylindrical "barbecue" basket. The basket is filled with natural rock that varies from 2.5 to 7.5 cm (1-3 in) in diameter. The surface area available for colonization is dependent on the substrate used in the basket.

Recommendations

The overall sampling design for artificial substrates is similar to the design for natural substrates described above. Artificial substrates should be used when it is too difficult to sample natural substrate or when there is too much variability in habitat (e.g., substrate) for matching site conditions. However, artificial substrates may not adequately assess contaminants associated with bottom sediments. The following is a series of recommendations for using artificial substrates (APHA, 1989; Klemm et al., 1990; ASTM, 1992):

- basket samplers are preferred over multi-plate samplers
- the sites selected should be as similar as possible to reduce variability
- in shallow streams, artificial substrates should be placed near (i.e., within 1 m) but not on the natural substrate; in deeper waters, the substrates may be more effective if suspended in the euphotic zone (i.e., the zone in which light penetration permits algal growth; usually within 1-5 m of the surface)
- substrates should be retrieved using a 180-250 Fm mesh net to prevent loss of invertebrates
- substrates should remain in place for a minimum of 6 to 8 weeks
- more substrates should be placed into a system than needed due to loss of substrates during the colonization period
- sampling should be done in the spring, late summer or fall depending on local conditions.

5.2 Field Procedures

Field procedures include all activities from the point of collecting a sample to the point where it is received by the laboratory for processing.

5.2.1 Sample Collection

Benthic invertebrate samples should be collected using predetermined procedures which are outlined in the sampling design. Procedures for the actual collection of samples using a particular device should follow standard methods (APHA, 1989; Klemm et al., 1990; ASTM, 1992).

For grab samples, it is important to have specific acceptance criteria for the sample collected. After the sampler has been retrieved from the bottom, the sediment should be inspected carefully before being accepted. The following acceptability criteria should be satisfied:

- overlying water is present (indicates minimal leakage)
- the sediment surface is relatively flat (indicates minimal disturbance or winnowing)
- the entire surface of the sample is included in the sampler

- the sample should have preestablished penetration depths.

If the sample does not meet these criteria, then it should be rejected.

For stream-net samplers, acceptability criteria are more related to the sampling process itself. For example, each sample should be collected using identical methods. This could include an established depth of sampler penetration and time for collecting each sample.

5.2.2 Sieving

If the sample collected in the field is large in volume, as is typically the case with grab samples, then the sample should be sieved in the field. The mesh size of the sieve (or the minimum mesh size if a series of sieves is used) should not exceed the recommended 180-250 Fm range, unless the objectives of the study allow for mesh sizes.

5.2.3 Containers

Sample containers must:

- be large enough so that the sample takes up no more than 50 percent of the container volume, with the remainder of the space allocated for preservative
- be sturdy for routine handling and transportation
- be leak-proof
- have physical and chemical properties that are not affected by the fixative/preservative
- comply with regulations concerning the transportation of dangerous good.

The recommended type of sample container is a plastic jar. Leak-proof heavy plastic bags may take up less room than jars if shipping or storage costs are a concern. Glass jars are not recommended for field use because they may break.

5.2.4 Fixative and Preservative

The recommended general-purpose fixative is a 10% formalin solution, which is both a preservative and a fixative. If the sample contains large amounts of organic material or numerous invertebrates, 20% formalin may be required. Some organisms may need to be relaxed before fixation, to prevent gross distortion or damage that renders them difficult or impossible to identify, although this is primarily a concern in marine studies. The length of time that specimens are kept

in formalin depends on the taxonomic group(s) in question. For instance, mollusc valves may be decalcified if exposed for long periods of time. The formalin solution should be buffered to reduce acidity, which causes the decalcification of molluscs. Ideally, pH should be at least 8.2. Borax (sodium borate) should be used as the buffer because other buffering agents may hinder identification by leaving a residue on body tissues. Most investigators prefer to replace the formalin with 70% alcohol or isopropanol after a week for long-term storage.

Many kinds of fixatives and preservatives are covered by regulations concerning the transportation of dangerous goods and by health and WHMIS (Workplace Hazardous Materials Information System) regulations. These need to be addressed during the planning of studies. Finally, various regulations and laws govern the disposal of chemicals such as formalin; even when these do not apply, the disposal of substances should be done in an environmentally sound and acceptable manner. Formalin can also be re-used if it is filtered through a sieve or cloth when samples are removed for transfer to alcohol or for sorting and identification. Alternatively, the used formalin can be stored in a large container, and solid material will settle out. The used formalin can be fortified (i.e., restored to 10% or greater concentrations) by adding concentrated formalin.

5.2.5 Staining Samples

Staining can be an aid to sorting, but its usefulness is largely a matter of individual preference. Stains should be mixed with formalin several days ahead of use. The sample may be stained either when collected, or later with a colouring agent to facilitate sorting. The most common stain used is rose bengal (Resh and McElravy, 1993). A rose bengal concentration of 4 g/L of concentrated formalin is commonly used; however, the exact concentration used may depend on the organic content of the sample.

5.2.6 Sample Labelling

Each sample should have two labels: one inside the container and one applied to the outside (not on the lid). These labels should, at the very least, include the site location, sample number and date. Material for labels should be appropriate for the fixative/preservative and container being used. Generally, water-proof paper and indelible ink are recommended.

A field log should be maintained that includes site and sample numbers and other pertinent data:

- latitude and longitude (or other numerical coordinate system) of each site as well as descriptive location name
- date
- local time (24-hr clock)
- names of crew members

- collector
- habitat descriptors
- collecting or sampling method
- sieving methods and mesh size
- other information (e.g., weather, stream flow).

5.3 Quality Assurance/Quality Control (QA/QC)

QA/QC procedures should be outlined during the development of the study plan and should be followed precisely to maintain high data quality. While a QA/QC plan for field procedures can have many components, some of the main procedures are as follows:

- all personnel involved in field procedures should have appropriate education and/or training
- sampling methods need to be consistent throughout study
- samples must be correctly collected (e.g., for Ponar grab - enough penetration, not washed out or overfilled, relatively intact, undisturbed sediment)
- sampling equipment must be appropriate for the particular habitat, properly cleaned, and with appropriate documentation
- all samples must have appropriate labelling
- samples must contain appropriate preservative/fixative
- personnel must maintain detailed field notes in ink in a bound, paginated notebook
- personnel must use chain-of-custody forms and custody seals
- personnel must follow appropriate and safe shipping and storage methods.

5.4 References Cited

APHA (American Public Health Association). 1989. Standard Methods for the Examination of Water and Wastewater. 17th edition. American Public Health Association, American Water Works Association and Water Pollution Control Federation.

ASTM (American Society for Testing Materials). 1992. Annual Book of ASTM Standards. Section 11: Water and environmental technology, Volume 11.04. American Society for Testing and Materials. Philadelphia, PA.

Doeg, T. and P.S. Lake. 1981. A technique for assessing the composition and density of the macroinvertebrate fauna of large stones. Hydrobiologia 80: 3-6.

Klemm, D.J., P.A. Lewis, F. Fulk and J.M. Lazorchak. 1990. Macroinvertebrate field and laboratory methods for evaluating the biological integrity of surface waters. EPA 600/4-90/030. U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Cincinnati, OH.

Scrimgeour, G.J., J.M. Culp and N.E. Glozier. 1993. An improved technique for sampling lotic invertebrates. Hydrobiologia (in press).

Wrona, F.J., P. Calow, I. Ford, D.J. Baird and L. Maltby. 1986. Estimating the abundance of stone-dwelling organisms: a new method. Can. J. Fish. Aquat. Sci. 43: 2025-2035.

SAMPLE PROCESSING

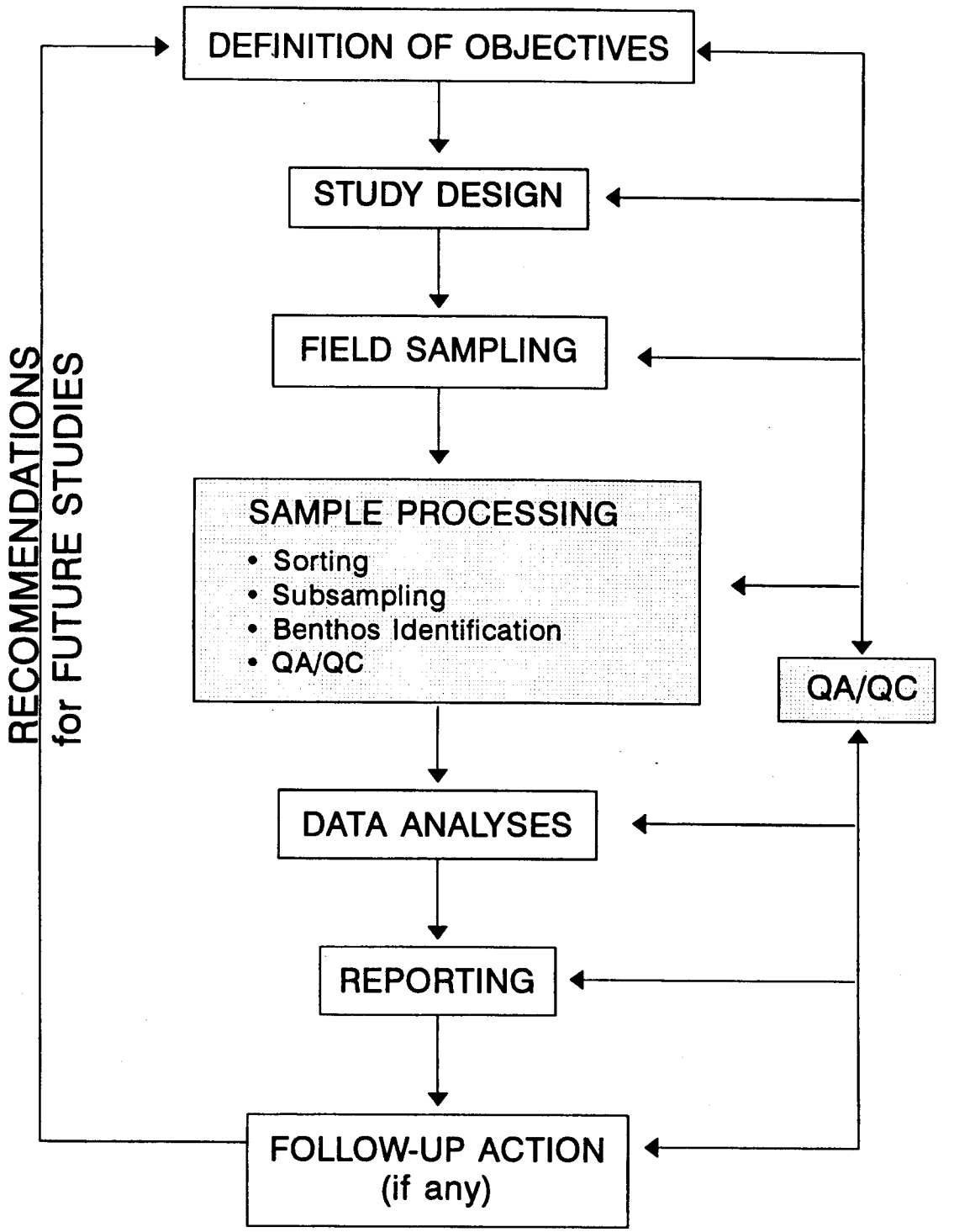

Figure 6. The key elements of a benthos monitoring study. Specific components of sample processing have been highlighted.

6.0 SAMPLE PROCESSING

This section reviews the processing of samples from the point of arrival at the laboratory through to data analyses.

6.1 Sorting

Sample sorting refers to the process of removing invertebrates from other sample material. Experienced sorters will also separate the recovered invertebrates into broad taxonomic groups to reduce the time required for identifications. The first step is to wash the sample to remove the preservative, and the preservative should be filtered, saved and re-used if possible. This is particularly important if samples have been stored in formalin. Samples should be washed using a sieve with a mesh size less than or equal to that specified in the sampling plan. Sorting should be performed by placing a small quantity of sample in a gridded petri dish and viewing it under a dissecting microscope. Each petri dish should be sorted twice to ensure that all organisms have been removed. The process is repeated until the entire sample has been sorted. The residue (i.e., detritus and other non-invertebrate material) should be saved so that the recovery efficiency can be checked in selected samples (see below). If possible, each sample should be sorted by a single individual to reduce additional sorting error. Records should be kept documenting the specific samples sorted by each sorter.

Sorting samples accounts for a significant portion of the labour involved in sample processing, and costs can be reduced by decreasing sorting time. Sorting time can be reduced by using sieves, stains or flotation techniques, or by subsampling (Section 6.2). Repeated sieving of the sample will remove fine particles such as silt or clay, but will not remove coarser material such as aquatic mosses or large detritus. Therefore, sieving is most effective for samples from lentic or depositional areas with a fine substrate (Resh and McElravy, 1993). Stains such as rose bengal are used to facilitate sorting by making invertebrates more visible. Samples may be stained when collected (Section 5.2.5). Flotation methods use solutions such as sucrose or calcium chloride which are denser than water or the sample preservative (Hynes, 1970; Klemm et al., 1990). If the sample is placed in a container such as an enamel pan and the sucrose or other dense solution added, most invertebrates will float to the top. These invertebrates are easily removed. However, the remainder of the sample must still be sorted using a microscope, as heavier organisms such as molluscs or caddis flies with stone cases will not float, and some organisms may remain attached to debris such as aquatic mosses.

Regardless of the sorting procedures used, recovery efficiency must be checked and must meet certain standards. The residue from 10-20% of the samples should be re-sorted, and the number of additional invertebrates recovered expressed as a percentage of the total. The consensus among workshop participants was that 95% recovery in initial sorts should be an achievable standard, and that if this standard were not met, all samples should be re-sorted. In samples in which one or a few invertebrates represent >5% of the total abundance, a standard could specify the maximum number of invertebrates which could be missed on the initial sort. For example, recovery of 9 of 10 invertebrates from a sample might be considered acceptable, even though <95% of the invertebrates are recovered.

6.2 Subsampling

Whenever possible, entire samples should be sorted and subsampling avoided. Techniques such as sieving, staining or flotation can reduce sorting time by up to 50% (Resh and McElravy, 1993), and may therefore be as effective as subsampling. However, when samples are extremely large in volume, as is typical with grab samples from depositional areas, subsampling may be required. The entire sample or some fraction of the sample (e.g., the invertebrates passing through a 500Fm mesh) may be subsampled. Subsampling is generally used to reduce sorting time, but may also be used to reduce the time required for identification. For example, Chironomidae or Oligochaeta may be subsampled after removal from the sample, if they are abundant and must be mounted on slides for identification.

Subsampling devices and methods for benthos samples are described in Hynes (1970) and Klemm et al. (1990). Methods and devices used for subsampling other organisms such as zooplankton or fish larvae could also be adapted for use with benthic invertebrates. Subsamples are usually taken by placing the sample in a gridded pan and sorting or identifying the invertebrates in randomly selected grids, or by mixing the sample in a large volume of some solution and sorting or identifying the invertebrates in one or more aliquots of the solution. Subsampling is usually conducted on preserved rather than live invertebrates, because preserved specimens are less susceptible to damage (e.g., when the sample is mixed to ensure an even distribution) and because live specimens are mobile and their movement will bias subsamples in grids or in aliquots.

Regardless of the subsampling method used, the following criteria must be met:

1) The fraction subsampled must be a known percentage of the total sample.

2) The subsample(s) must be representative of the entire sample (i.e., the subsample should not be biased towards or against certain taxa).

3) The variance associated with subsampling should be small relative to the variance among replicate samples. If this criterion is not met, then subsampling will reduce the power of any statistical tests. Additional replicates will have to be collected to provide the same power achieved by sorting entire samples, and the time and cost associated with subsampling and sorting these additional replicates may exceed any savings gained from subsampling.

4) Subsampling methods must be easy to use, and substantially reduce the time required for sorting or identification. The best methods would be effective for a variety of sample types and substrates, so that different methods would not be required for each situation.

The precision and accuracy of the subsampling method used should be assessed initially by comparing selected subsamples to the total sample. One method of tracking subsample performance is to split the sample into many parts, then use several of these parts as a subsample. For example, the sample could be split into 32 parts, the invertebrates in 8 of these parts identified and counted, and abundances totalled to provide a subsample of one-quarter (8/32) of the

original sample. The variance in abundance among the eight (1/32) parts can also provide useful information. First, if the variance of total abundance (or abundance of any taxa) is much greater than the mean, then the splitting procedure is not distributing the invertebrates randomly or evenly among the various fractions. Second, the standard error (SE, or square root of the {variance/8}), multiplied by the number of fractions (i.e., 32), is the standard deviation (SD) attributable to subsampling. This subsampling SD can be compared to the SD among whole replicate samples to determine if subsampling adds significant variance.

The fraction subsampled depends in part on the volume and nature of the initial sample, but Klemm et al. (1990) suggest that the subsample should represent at least one-quarter of the original sample, and contain at least 100 organisms. It may not be possible to follow these suggestions in all cases; for example, if large volumes of lake sediments with few organisms are subsampled or if all samples are subsampled to standardize methods (see below). However, the suggestions are based on practical experience, and investigators should document that departures from these suggestions do not introduce substantial bias or imprecision.

Many subsampling methods do provide relatively precise and accurate estimates of total abundance and the abundance of common taxa, and the variance among subsamples is often much smaller than the variance among field replicates. The cost of collecting additional replicate samples in the field is often trivial, so statistical power per unit cost can be increased in many studies by collecting more field replicates and subsampling all samples (Resh and McElravy, 1993). The major disadvantage of subsampling is that abundances or even presence/absence of rare taxa are imprecisely estimated (Wrona et al., 1982). There are sound reasons to eliminate rare taxa from statistical analyses (Section 7.0), but subsampling will also affect estimates of richness (number of taxa), usually in unknown or unpredictable ways. Richness for the total sample cannot be estimated by dividing richness in a subsample by the fraction subsampled (i.e., in the same manner as abundance is estimated). Most common taxa will be present in the subsample, but the rarer taxa may not be. The best solution is to subsample all samples in the same manner, and compare richness per subsample without attempting to estimate richness per sample. If other approaches are used, such as removing common taxa (e.g., Chironomidae) from a subsample only but removing other taxa from the entire sample, investigators should document the effects of the approach on estimates of richness.

6.3 Benthos Identification

Depending on the objective(s) of a study, the identification of benthic invertebrates can range from the phylum down to the species level. The workshop recommended that, while it is ideal to identify all organisms to the lowest taxonomic level possible, establishing a set taxonomic level for all studies is undesirable. Rather, the levels need to be set and rationalized in each case to meet the study objective(s). The arguments for and against using the lowest taxonomic level possible are discussed in Resh and McElravy (1993), and their summary adequately captures the conflicting points of view offered at the workshop.

For taxonomic work, workshop participants provided the following recommendations:

- identifications should be verified by an expert in the taxonomic group of interest
- persons who carry out the identifications should be named, with appropriate details of qualifications
- literature and taxonomic keys used for benthos identification should be referenced
- details of both reference and voucher specimens/collections should be given (institutions or agencies holding the specimens; catalogue numbers of the specimens; etc.).

The workshop participants identified a need to develop an approved list of experts in taxonomy and identification of freshwater organisms so that there could be some consistency across studies. This could eventually evolve into a system of accreditation for identifying freshwater organisms (e.g., through a standard course or courses in identification). The North American Benthological Society maintains lists of experts who will provide verification services (often free of charge), and the lists are available from the Society or in the Society Bulletin.

6.3.1 Reference/Voucher Collections

"Reference" collections and specimens are used to identify specimens sampled in a benthos study. They complement the use of literature and keys.

Several principles apply to the use of such materials in benthos monitoring studies:

- identification of private collections should be independently verified
- the authorities who identified or verified the specimens should be named, their affiliations given, and the dates of identification/verification indicated
- where available, the same information for museum collections should be given
- the location of the reference material used, and its catalogue or reference numbers, should be listed.

"Voucher collections" (or specimens) are representative collections from field research or surveys, preserved to permit independent verification of results and to allow further study. The workshop recommended the deposition of voucher collections for all studies. The nature of a voucher collection and details of its deposition (agency or institution name; catalog or reference numbers; etc.) should be indicated in each final or data report.

Each sample should have two labels: one inside the container and one written or applied to the outside. These labels must, at the very least, include the site and sample number. Material for labels should be appropriate for the fixative/preservative and container being used.

6.4 Quality Assurance/Quality Control (QA/QC)

QA/QC for sample processing involves the following elements (many of which were mentioned above):

- Appropriate procedures should be used for handling, sorting, subsampling, preserving, identifying and storing benthos samples. Departures from these procedures should be justified and documented.

- Sorting and re-sorting should be done by trained personnel using a microscope and appropriate procedures and should be fully documented. Standards for recovery efficiency should be established, and 10-20% of samples should be re-sorted to ensure that these standards are met.

- The precision and accuracy of subsampling methods, especially with respect to estimates of abundance and richness (number of taxa), should be documented.

- Identifications should only be done by reliable identifiers, and their identifications should be checked by knowledgable experts. Taxonomic references used for identification must be provided.

- All studies should develop a reference collection which is accessible by all individuals involved in the study. The collection should be maintained for the duration of the study, and preferably for several years after.

- Voucher collections should be deposited with museums or research institutes.

6.5 References Cited

Hynes, H.B.N. 1970. The Ecology of Running Waters. University of Toronto Press, Toronto, Ont. 555 pp.

Klemm, D.J., P.A. Lewis, F. Fulk and J.M. Lazorchak. 1990. Macroinvertebrate field and laboratory methods for evaluating the biological integrity of surface waters. EPA 600/4-90/030. U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Cincinnati, OH.

Resh, V.H. and E.P. McElravy. 1993. Contemporary quantitative approaches to biomonitoring using benthic macroinvertebrates. pp. 159-194. In: Freshwater Biomonitoring and Benthic Macroinvertebrates. D.M. Rosenberg and V.H. Resh (eds). Chapman and Hall, New York, NY.

Wrona, F.J., J.M. Culp and R.W. Davies. 1982. Macroinvertebrate subsampling: a simplified apparatus and approach. Can. J. Fish. Aquat. Sci. 39: 1051-1054.

DATA ANALYSES

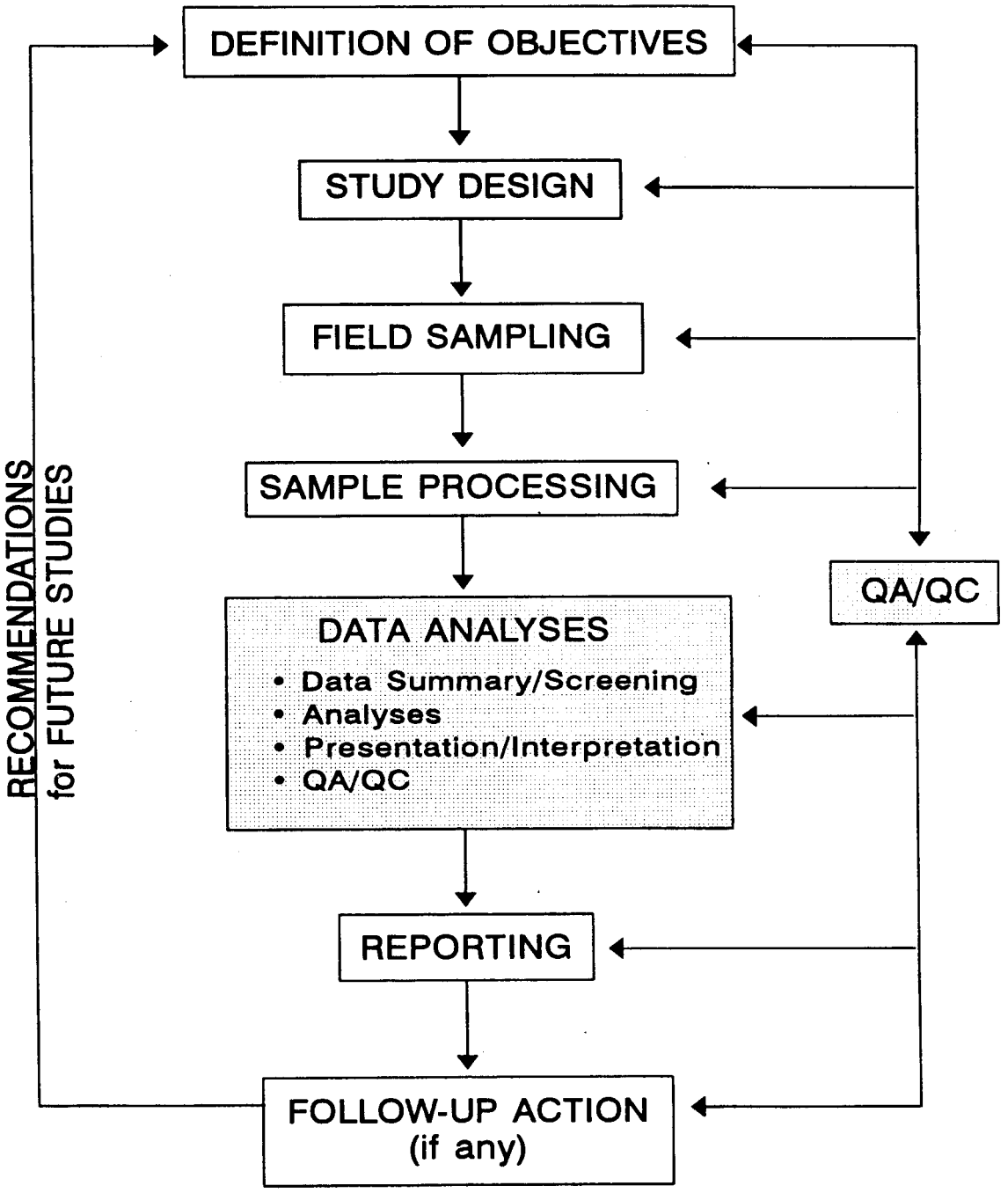

Figure 7. The key elements of a benthos monitoring study. Specific components of data analyses have been highlighted.

7.0 DATA ANALYSES

Data from benthos monitoring studies consists of matrices of taxon abundances or biomass for several sites and/or times. Most analyses of these data examine distance or similarity matrices. The distance matrices consist of some measure of the distance between pairs of sites, times, or site-time combinations. The objective of the analyses is to determine if patterns in these distance matrices correspond with patterns expected if impacts were present. Analyses are almost always multivariate, with the taxa, or less frequently the sites and/or times, as the variables. Investigators may also compare distance matrices based on taxon abundances with matrices based on habitat variables. These habitat variables could include physical factors such as depth or substrate type, and may also include impact-related variables such as contaminant concentrations.

There are four components to data analysis:

- data summary, screening and manipulation
- specific data analysis methods
- data interpretation and presentation
- Quality Assurance/Quality Control (QA/QC).

Green (1979) and Gauch (1982) provide good general reviews of methods suitable for analyzing multivariate community data. Burd et al. (1990), Resh and McElravy (1993), Norris and Georges (1993), and various federal, provincial and state manuals (e.g., Klemm et al., 1990) provide methods specifically for benthos community studies. These reviews and manuals offer a broad range of methods for data analyses, and recommendations vary among authors. For example, the methods recommended in the workshop differed from those recommended in the draft EEM guidelines (Environment Canada, 1991). The choice of methods in the workshop was guided by two important criteria:

- Methods should test specific impact-related hypotheses which are suggested by the study design. Methods designed for general data "snooping" or "fishing expeditions" should be of secondary importance
- Methods should objectively identify relationships among variables. Derived variables such as many of the indices commonly used in benthos monitoring studies make *a priori* assumptions about relationships among variables and should be used for presentation rather than for analysis.

7.1 Data Summary, Screening and Manipulation

The first step in data analyses is to check data entry for transcription errors, as failure to identify these errors can invalidate the analyses. Assuming that data have been entered correctly, the following steps are recommended:

- data summary and initial screening
- validating normality and homogeneity of variance assumptions
- data transformation
- data reduction.

Green (1979) and most statistical texts (e.g., Winer, 1971; Snedecor and Cochran, 1980; Sokal and Rohlf, 1981; Tabachnik and Fidell, 1983) review procedures for data summary, screening and manipulation.

7.1.1 Summary and Initial Screening

Before starting any analyses, data should be summarized with means and standard deviations (SD) calculated. These summary statistics can be calculated over the entire data set (grand means) and for each cell (site and/or time). The summary statistics are usually required for presentation, and can also be used in the screening and manipulation steps discussed below. For example, the most abundant taxa can be identified from grand means. Relationships between cell means and SD can be used to check for homogeneity of variances and to suggest appropriate transformations. There are also visual screening techniques such as box-and-whisker plots, normal probability plots and stem-and-leaf diagrams, which are useful for identifying extreme values (true outliers or data entry errors). Most statistical software packages provide data summary modules which include summary statistics and graphics.

Another useful summary step in multivariate analyses is to calculate the correlation matrix among variables. In many studies, these correlations would be between taxon abundances. However, correlations between physical or chemical variables can also be calculated. Scatter plots should also be examined to determine if relationships among variables are linear, an important requirement for many multivariate analyses. The correlations and scatter plots are useful for identifying outliers or data entry errors, suggesting transformations to linearize relationships, and identifying potentially redundant variables if the number of variables must be reduced. There is also no point conducting most multivariate analyses unless there are strong correlations among variables (Green, 1979).

7.1.2 Normality and Homogeneity of Variance

Parametric methods such as analyses of variance (ANOVA) and least-squares regression require that data (actually residuals) are normally distributed and that variances are homogeneous (equal) among cells. There are two approaches to determining whether these assumptions are met: formal testing, and inspection of residuals. Both approaches pose problems for multivariate tests, as it is difficult to determine if multivariate normality exists, and if covariance matrices are equal across cells. Investigators should also note that parametric procedures are reasonably robust even when these two assumptions are violated provided that the data set is large (>50 observations or replicates) (Tabachnik and Fidell, 1983).

Formal Testing

There are various tests for normality, such as the Shapiro-Wilks Test and the Kolmogorov-Smirnov Test (Klemm et al., 1990), which compare the distribution of residuals (deviations from cell means or regressions) with a normal distribution. Similarly, Bartlett's Test and Levene's Test determine whether variances are homogeneous among cells (Snedecor and Cochran, 1980). These tests will indicate whether there are significant departures from normality or homogeneity of variances, but do not provide any indication of the causes and consequences of violating these assumptions. Violations can occur when the departures from normality or homogeneity of variances are systematic, such as when the underlying distribution is log-normal. In that case, the assumptions can be met if variables are transformed. Distributions may also be bi- or even multi-modal, which suggests that some confounding factor is present and should be accounted for or removed. Violations can also occur when outliers are present, as the variance of cells containing the outliers will be inflated, and the distributions in these cells will be skewed towards the outlier. There are formal procedures available for detecting "significant" outliers, but most investigators are reluctant to remove outliers unless they represent obvious sampling, measurement or data entry errors (Grubbs, 1969; Green, 1979). Usually the best procedure is to run analyses with and without the suspected outlier(s) to see if they actually affect results and conclusions.

Formal tests for normality, homogeneity of variances, and outliers are not commonly used in ecological studies. Some authors (e.g., Day and Quinn, 1989) specifically recommend that the formal tests should not be used.

Examination of Residuals

Draper and Smith (1981) review various methods for examining residuals, particularly residuals from regressions. Most statistical software packages also provide modules for examination of residuals. These methods are usually graphical, although there are diagnostic statistics available as well. The primary advantage of these methods, compared to formal tests, is that they can identify the cause of the violations. For example, normal probability plots, box-and-whisker plots, and plots of residuals *versus* predicted values can be used to identify outliers and to suggest transformations. Using these methods can be subjective, and some experience is required to identify patterns suggesting specific transformations.

7.1.3 Transformations

Formal procedures such as Taylor's Power Law, which examines the relationship between cell means and variances, can be used to determine appropriate transformations to normalize data or homogenize variances (Green, 1979). Selection of appropriate transformations for linearizing relationships among variables is more subjective, although scatter plots of the raw data and various plots of residuals are useful. In most cases, a logarithmic transformation will be adequate, and more comprehensible and biologically meaningful than exotic transformations such as the fourth root. If zeroes are present, a $\log(x+c)$ transformation should be used, where c is the lowest non-zero value in the data set.

Occasionally, relative abundances (percent of total) rather than raw abundances are analyzed. The arcsine transformation is often used for percentages (Sokal and Rohlf, 1981), but will not equalize variances unless total abundances are similar for all cells. One possible solution is to use the total abundance as a covariate (X) in ANCOVA with the individual taxon abundances as dependent variables (Y), but the relationship will be non-linear if taxa are rare (i.e., close to 0% of the total) or dominant (i.e., close to 100%) in several samples. Non-parametric methods or transformation to ranks or rankits (see below) will usually be effective for relative abundance.

Non-parametric tests are really a special case of transformation. Transformation of ranks to rankits or probits will normalize the entire data set, but not necessarily the residuals (Green, 1979). In large data sets (> 50 observations), the data can usually be transformed to ranks, and parametric tests conducted on the ranks (Conover, 1980). There are also non-parametric equivalents to ANOVA and other parametric procedures. In general, transformation to rankits/probits or non-parametric tests are effective when distributions are non-normal and/or to reduce the influence of outliers, but are not effective when variances are unequal (Day and Quinn, 1989). There are also non-parametric cluster and ordination methods available (Gauch, 1982; Burd et al., 1990).

A final type of transformation is the use of presence/absence data. Parametric procedures are reasonably robust when presence/absence data are used if total sample size is >50 (Green, 1979). Most ordination and cluster methods can use presence/absence as well as continuous data. Presence/absence data are useful when abundances are low but variable, or when there is concern about the elimination of rare but sensitive taxa.

7.1.4 Data Reduction

There are limitations or restrictions on the number of variables which can be used for parametric methods, such as multivariate analyses of variance (MANOVA) or analyses of covariance (ANCOVA) (Green, 1979). For MANOVA, the total number of observations (N) must exceed the number of variables (p) plus the number of cells (i.e., sites, times, or site-time combinations) (q), or there will be no error degrees of freedom for conducting tests. More generally, investigators should be wary of analyzing any data set where the number of observations (replicates; N) is less than five times the number of variables (p). Tabachnik and Fidell (1983) suggest that N=10 to 20p might be more reasonable for hypothesis testing procedures.

Descriptive and classification methods such as cluster analysis (see Section 7.2.2) can theoretically accommodate an infinite number of variables (usually species), although most computer programs will have limitations on the number of variables which can be analyzed in practice. Gauch (1982) recommends dropping rare species, because they usually either contribute nothing to the analysis or behave as outliers.

The number of variables can be reduced by either dropping or combining variables. Regardless of which method is used, objective rather than arbitrary rules or procedures should be used. If variables are taxon abundances, the most obvious method for dropping variables is to eliminate rare taxa. For example, one could select the p most abundant taxa, with p set as some fraction of N. Alternatively, one could select all taxa whose abundance was greater than some percentage of the total (e.g., 1 or 5%), or select the most abundant taxa until the cumulative percentage reached some specified value (e.g., 75 or 90% of the total). In the case of habitat variables, one could use a cost-power function to choose the variables providing the most statistical power per unit cost (e.g., Ferraro et al., 1989). Correlation matrices can be used to identify groups of highly correlated (redundant) variables, from which only one or a few would be selected for analyses (Green, 1979).

Variables can be combined in arbitrary ways, such as when ratios or indices are calculated. This approach should be avoided (see Section 7.2.2). Taxon abundances can be combined by pooling abundances at higher taxonomic levels. The disadvantage of this procedure is that all species within a higher taxon may not be similar, especially in terms of their sensitivity and response to pollution or other impacts. Perhaps a useful set of rules would be to use higher taxa only when:

- no information on tolerance or sensitivity is available for lower taxonomic levels, or
- all or most taxa at lower levels are similar ecologically relative to the difference among higher taxa, or
- abundances of all taxa within a higher taxon are positively correlated.

More formal and objective techniques for combining variables are described in Green (1979); these include many of the ordination methods described in Section 7.2.2. All multivariate techniques combine variables, because they all calculate measures of distance which are combinations of the original variables. Methods such as principal components

analysis (PCA) are especially effective for combining variables representing habitat parameters or chemical concentrations for use in regressions or ANCOVA. For example, one or two principal components (PC) might account for most of the variance in a data set of 15 contaminant concentrations, as these concentrations are usually positively correlated.

7.2 Data Analyses

7.2.1 Primary Methods

The primary methods recommended for analysis of benthos monitoring data sets are ANOVA, MANOVA, and ANCOVA with planned comparisons (contrasts). The variables analyzed should be simple measures such as total abundance and richness (number of taxa per sample), and individual taxon abundances. Various methods for removing or testing the effects of habitat or other factors are briefly discussed.

Univariate Analyses (ANOVA, ANCOVA)

In benthos monitoring studies, ANOVA is used to compare simple community measures such as total abundance and richness among sites, times, or both. Taxon abundances may also be analyzed in univariate analyses if only one or two taxa are of interest; many statistical packages will also provide univariate tests as part of MANOVA output. Although univariate analyses are much simpler, and less time-consuming, to conduct than multivariate analyses, their importance should not be ignored. Variables such as total abundance, richness, and the abundances of dominant taxa convey considerable information, and are usually sensitive to impacts (Resh and Jackson, 1993). There are numerous cases in which these variables adequately identified potential effects or impacts. These variables are also useful for conducting power and power per unit cost analyses, comparing sampling designs or sampling methods, and evaluating subsampling procedures, as similar analyses for the multivariate case are difficult to conduct.

Normally, investigators are interested in testing specific differences among sites or times rather than simply testing the overall site or time effect. The most commonly used procedures make all possible pair-wise comparisons among means, such as multiple range tests (see Day and Quinn, 1989, for a review). These procedures are inefficient, and have low power in terms of detecting differences of interest, because probability values must be adjusted to account for the large number of comparisons made. Making pair-wise comparisons, or *post hoc* comparisons suggested by the data, is data "snooping" and does not take advantage of aspects of the study design such as the location of sites at various distances from a discharge (Hoke et al., 1990).

In most monitoring studies, only a few impact-related comparisons are of interest, and planned comparisons (e.g., orthogonal contrasts) should be the method of choice (Sokal and Rohlf, 1981; Hoke et al., 1990). For example, suppose a study includes two reference sites, three near-field sites, and two far-field sites. There are 21 possible pair-wise comparisons of sites. However, two differences representing independent contrasts are especially relevant for assessing potential impacts. First, the difference between the reference sites and the pooled near- and far-field sites provides an indication of the overall downstream impact. Second, the difference between the near- and far-field sites indicates whether

impacts extend to the far-field sites. More contrasts could be made (up to a total of six). For example, a contrast comparing two reference sites would indicate whether non-impact (i.e., "natural") differences among sites were as large as differences between the reference and impact sites. One could also use contrasts to test for a linear or quadratic relationship between any variable(s) and effluent concentration or distance from a source. The existence of a strong dose-response relationship would be difficult to attribute to some natural factor and would strengthen the evidence for an impact from the source.

ANCOVA can be used to remove or test the effects of habitat or other variables, provided that only one or a few covariates are used. ANCOVA can also be used to analyze regressions of richness on abundance, as there is often a positive relationship between the two variables. One could analyze evenness (richness divided by abundance) instead, but doing so assumes a linear relationship between the two variables which passes through the origin. In most cases, the true relationship is curvilinear, approaching an asymptote determined by the number of taxa actually present in the study area (i.e., similar to species-area curves).

Multivariate Analyses (MANOVA)

MANOVA is analogous to ANOVA, except that more than one variable is analyzed. In MANOVA, sites and/or times are factors, and contrasts can be used to test impact-related hypotheses. Note that it can be difficult to adapt procedures such as multiple range tests to MANOVA, another reason why contrasts are the method of choice.

MANOVA provides multivariate tests of significance for factors (e.g., sites) or contrasts. There are a number of test statistics available, and probability values will usually differ slightly among these statistics except for single degree-of-freedom contrasts (Wilkinson, 1990). MANOVA also provides canonical vectors or discriminant functions; the number of vectors is equal to the degrees of freedom for the factor or contrast. These vectors are linear combinations of the original variables, with each variable multiplied by a coefficient or weight. The values of the vectors for individual replicates are referred to as canonical scores. The vectors provide the maximum possible separation of factor levels in multivariate space, or maximize the difference associated with a contrast. The vectors provided from MANOVA are similar to some indices, in that taxon abundances are assigned weights for both indices and vectors. In many cases, vectors closely resemble one or more indices, and the index values can be used for presentation. Thus, MANOVA can be useful in objectively identifying which of the many available indices are useful in describing impact-related changes.

Interpretation of vectors is best illustrated by example. Table 3 provides standardized coefficients (SC) and loadings for a contrast comparing near- and far-field sites in a recent study conducted by Gibbons et al. (1992). The loadings, or correlations between vector scores and the original variables (taxon abundances) were all positive, indicating that the vector reflects differences in total abundance. This conclusion was verified by analyses of total abundance. The SC are the weights or coefficients for each taxon divided by their standard deviations within groups. Standardization removes the effects of variance within groups, which is of no value in separating groups. Some SC were positive, and others negative, which is usually an indication that there are differences in relative (% of total) as well as raw abundances. The EPT (Ephemeroptera, Plecoptera, Trichoptera) taxa (*Ephemerella, Isoperla, Hydropsyche*) had positive SC whereas the Oligochaeta and Chironomidae had negative SC. These SC indicate that the relative abundance of the EPT taxa was low

when that of the Chironomidae and Oligochaeta was high, and *vice versa*. In this particular study, abundances of all taxa, but especially the EPT taxa, increased immediately downstream of the discharge relative to baseline conditions.

Plots of canonical scores are another useful aid to interpretation. If several vectors are generated, such as for a factor (e.g., sites), two-dimensional plots of the first and second vectors are often sufficient to separate levels of the factors. For a single degree-of-freedom contrast, such as a comparison between near- and far-field sites, the plot would be one-dimensional. These plots indicate the extent of separation of groups, which is probably the best expression of effect size in multivariate studies. Many statistical packages will provide the percentage of correct classification of groups, which is another way to measure separation or effect size. In the case of vectors generated by the general comparison among sites, the percentage of correct classification would be the percentage of replicates correctly classified (i.e., assigned to the appropriate site). Comparable measures can be generated for contrasts; interactions (e.g., between time and site) may present more difficulties. For a contrast comparing near- and far-field sites, one could calculate the percentage of near-field replicates which were outside the range of values for far-field replicates or *vice versa*.

Adjustments for Habitat and Other Factors

The potentially confounding effects of habitat variables can be removed by using the original variables, or PC derived from them, as covariates in ANCOVA or MANCOVA. If the habitat variables are categorical, such as "riffle" and "pool", they can be used as factors in analyses or the two categories can be analyzed separately. Another method is to compare distance matrices derived from species abundances (i.e., community data) with matrices derived from habitat or other variables. Mantel's Test is the appropriate test for such comparisons, and Legendre and Fortin (1989) provide a good review of procedures with numerous examples. Mantel's Test calculates some measure of association between two matrices (e.g., the rank correlation between pair-wise distances from each matrix), and then compares the value of that measure with the distribution generated by randomly sampling the data to calculate the probability of obtaining a higher or lower value than that observed. Mantel's Test can be very useful, for example, to test whether community distance matrices are

Table 3. Standardized coefficients (SC) and loadings (correlations, *r*) for five dominant taxa in the Lesser Slave River for a vector generated by a contrast of near- *versus* far-field sites (from Gibbons et al., 1992).

Taxon	SC	*r*
Oligochaeta	-24	32
Chironomidae	-32	19
Hydropsyche	24	59
Isoperla	13	57
Ephemerella	100	92

NOTE: SC and *r* have been multiplied by 100.

similar to distance matrices based on contaminant concentrations. However, the test does suffer from the same problems as many of the secondary methods described below in that it cannot address specific impact-related hypotheses that may be independent of other dominant environmental gradients.

7.2.2 Secondary Methods and Variables

Secondary Methods

Secondary methods include ordination and cluster techniques, which do not directly address impact-related hypotheses. These techniques are reviewed extensively in Green (1979), Gauch (1982), and Burd et al. (1990), and will only be briefly considered here. These methods examine distance matrices calculated among sites or times ("Q" analysis), among species ("R" analysis), or both (e.g., reciprocal averaging and detrended correspondence analysis).

Ordination techniques plot sites or species on one or more axes, which are usually combinations of the original variables. These plots indicate the major groupings of sites or species, if any exist, and the degree of separation between them. Identification of groups is subjective. Cluster techniques group similar sites or species, based on the distances between them. Most cluster methods generate hierarchical clusters, with subgroups nested within larger groups. The investigator then subjectively determines how many "real" groups exist or, if replicates are present, uses randomization tests to determine which groups differ significantly.

Both ordination and cluster techniques are descriptive, data "snooping" methods. The techniques can only identify the major groupings and the dominant environmental gradient(s) present. For example, reciprocal averaging or PCA might identify groups of fast-water and slack-water sites, using axes which emphasize the abundance of sprawling and clinging taxa *versus* swimming and climbing taxa. Similarly, a cluster analysis might generate clusters of fast-water and slack-water sites. The presence of a strong environmental gradient such as current velocity might obscure, or be confounded with, any axes related to impacts, making it impossible to identify impact-related changes.

Although secondary methods are not suitable for specifically testing impact-related hypotheses, they can be very useful for reconnaissance or baseline studies and for data reduction. For example, if sites were grouped into fast-water and slack-water sites after a reconnaissance or baseline study, regardless of their position with respect to an existing or proposed discharge, then the effects of current velocity could be removed in the actual monitoring study by sampling only fast-water sites, using fast-water and slack-water sites as levels of a factor, or using current velocity as a covariate. The axes from ordinations can be used as covariates in ANCOVA or MANCOVA, if there are too many original variables. Investigators can also use ordination and cluster techniques to conduct "R" analyses which will identify groups of species (possible sub-assemblages) whose abundances are correlated. Then only one species in a group could be used in analyses, or species could be pooled into higher taxa or functional groups as suggested by the data.

Secondary Variables

Secondary variables are derived variables such as indices, which are constructed by combining original variables (taxon abundances, richness and total abundance) in some fashion. If the derived variables are objectively derived via the data reduction techniques described in Section 7.1.4, then it may be legitimate to use them as primary variables in ANOVA or MANOVA. However, investigators should be very cautious about using derived variables such as most indices, because they assume some relationship between the variables which may not exist, and because they may have unusual and undesirable statistical properties (Green, 1979). In many cases, especially when dealing with ratios such as evenness, it is better to use the original variables in ANCOVA.

Although indices and other derived variables may not be suitable for analyses, they may be very suitable for presentation. In most cases, there is some logic and meaning to the construction of indices and their values; the problem is to identify the most useful of the many indices available. In the example provided in Table 3, the SC for the MANOVA vector indicated that the relative abundances of the three EPT taxa were inversely related to the relative abundances of the Chironomidae and Oligochaeta. The results could be summarized effectively by presenting the relative abundances of the EPT taxa for each site and time, or by presenting the abundance of EPT divided by the abundances of Chironomidae and/or Oligochaeta. Both of these forms of presentation correspond with indices commonly used in the United States (Klemm et al., 1990; Resh and Jackson, 1993), and assume that the EPT taxa are sensitive and the Chironomidae and Oligochaeta tolerant. The latter assumption would be questionable, unless the dominant Chironomidae and Oligochaeta genera or species present were tolerant (i.e., there are many sensitive species within these large taxa).

7.3 Data Presentation and Interpretation

Some issues related to presentation have been discussed in preceding sections. These are:

- provide summaries (usually means ±SD) of the variables (untransformed or back-transformed)
- provide plots of canonical scores for replicates or sites
- provide percent correct classification or some similar measure, as well as the results of significance tests
- use indices suggested by MANOVA vectors to present and discuss results.

Graphical presentations of these and other data are often more effective than tables of numbers. For example, relative abundances of various taxa at each site or time are probably best summarized in a pie chart.

Data interpretation is simplified by testing specific impact-related hypotheses. The investigator does not have to construct an argument for the presence or absence of impacts from ordinations or clusters which only reflect the dominant environmental gradients or groupings present. The investigator can simply state that, for example, the

contrast testing for significant differences between near- and far-field, or between reference and downstream (impacted) sites was or was not significant, which does or does not suggest the presence of near-field and/or far-field impacts. As in any observational study or natural experiment, there is always the possibility that the differences were due to some unmeasured non-impact factor. However, use of habitat and other potentially confounding variables as factors or covariates will considerably reduce this possibility. Potential problems also arise if analyses (e.g., MANOVA) identify significant impact-related differences, which are not easily interpretable as positive or negative. For example, the vector from a MANOVA may indicate a positive association between intolerant and tolerant taxa, but negative associations within those two groups. In such a case, the investigator should probably describe the difference and indicate that it is impossible to classify a positive or negative impact.

A major issue, raised repeatedly in the workshop, was the definition of biologically (as opposed to statistically) significant effects. Two approaches, neither entirely satisfactory, have been or could be used:

1) Define a biologically significant effect in terms of effects on fish. The appeal of this approach is that aquatic environmental regulation in Canada depends heavily on the Federal Fisheries Act, and that regulators, managers and the general public are more familiar with, and probably more concerned about fish than benthos. Hanson and Leggett (1982) provide empirical relationships between benthos biomass and fish yield for lakes; similar relationships could be developed for streams and rivers. Thus, the effects of changes in benthic biomass on fish yield could theoretically be predicted, but the predictions would probably be too imprecise to be of value. Furthermore, variables such as total abundance and taxonomic richness, and changes in community structure, would be difficult to relate to fish yield. Any approach which attempts to translate effects on benthos into effects on fish adds uncertainty to estimates of effects, and still requires a definition of biologically significant effects for fish. Furthermore, emphasis on effects on fish implies that effects on benthos are not important in themselves. We doubt that Canadian regulators are prepared to allow unregulated discharge into streams and lakes wherever there are no fish.

2) Define biologically significant effects as specific effect magnitudes (e.g., 10% increase in biomass over reference). The effect magnitude chosen would be arbitrary to some degree, but regulators make similar potentially arbitrary decisions when setting permit limits or criteria based on laboratory toxicity tests. There are probably sufficient data on impacts on, and natural variation of, abundance, richness, and relative abundance of dominant taxa that some consensus could be reached on defining biologically significant changes for these variables. Defining biologically significant effects for multivariate metrics would be more difficult, and can probably be done only in narrative terms. The literature on recovery of disturbed benthic communities (reviewed by Yount and Niemi, 1990) could also be used to define biologically significant effects as irreversible changes or as changes that persist for some specified time (e.g., more than one life cycle) before recovery occurs.

The Capital Region District (CRD; city of Victoria, B.C., and surrounding communities) has defined guidelines for marine benthic communities which incorporate the second approach given above (CRD, 1992), and is one of the few agencies in Canada to do so. Effects on total abundance and richness are defined as differences of >25% between reference and impact stations, with the provision that variation in these variables should not be correlated with

an environmental gradient not related to the outfall. Effects on community structure are defined narratively. An effect is considered present if ordination separates stations nearest the outfall from those further away or reference stations, and if the ordination does not reflect an environmental gradient not related to the outfall. The effect sizes for abundance and richness were chosen because smaller changes are well within the bounds of natural variance among stations or replicates, and because the CRD (1992) uses similar effect sizes in their guidelines for sediment toxicity tests. The CRD will continue to evaluate and refine these guidelines in future monitoring studies.

There may be other approaches to defining biologically significant effects, but regardless of which approaches are chosen, regulators and scientists need to address the issue. The statistical procedures provided in this section can identify changes in benthic communities, and help separate natural changes from potential impacts from human activities, but cannot make regulatory decisions or value judgements.

7.4 Quality Assurance/Quality Control (QA/QC)

The major concerns with respect to QA/QC are:

- data verification and validity
- repeatability and robustness of the analyses
- rigour and defensibility of the analyses

The validity of data depends on quality control of sample collection, sorting, and identification procedures, which are discussed elsewhere. The screening techniques discussed in Section 7.1 can identify transcription errors and other "suspicious" data points.

Any analyses conducted should be repeatable, in that another investigator should be able to reproduce the analyses and results. The first step is to provide the raw data in an Appendix, and retain computer data files for some fixed period (e.g., one year) after the analyses are published in a report. If the data are to be compared to other data collected in the future as part of a long-term monitoring program, the data files should be maintained through the life of the project. The second step is to provide a clear description of the methods used for analyses. Investigators should also verify that the statistical packages used produce the same output and results as do other packages. Investigators should be wary of packages developed by private individuals for specific ordination and cluster analyses, unless these packages have been widely used and frequently "de-bugged" and revised. Commercial packages such as BMDP, SAS, SPSS, and SYSTAT, and packages developed by universities and research institutes are usually subjected to extensive verification and revision.

Analyses are robust if results and conclusions are similar regardless of whether outliers are included or excluded, transformations are used or not used, and specific taxa are included or excluded. Analyses can be conducted with and without specific replicates, sites, or variables to determine the influence these replicates, sites, and variables have on

conclusions. The results obtained using different taxonomic levels or different distance measures could also be compared. The objective is to ensure that the results are not a function of some manipulation or assumption made prior to or during analyses.

Analyses should be rigorous, in terms of using the best methods available which directly test hypotheses of interest. The effects of confounding factors should be estimated or eliminated, and every effort should be made to test and eliminate alternative non-impact hypotheses which might explain the results. These approaches should make the analyses defensible. Investigators should also indicate the rationale for choosing particular methods, and understand exactly what specific methods can and cannot do or provide. This is particularly important for complex multivariate methods, as there are many methods to choose from, and the choice made must be defensible.

7.5 References Cited

Burd, B.J., A. Nemec and R.O. Brinkhurst. 1990. The development and application of analytical methods in benthic marine infaunal studies. Adv. Mar. Biol. 26: 169-247.

Conover, W.J. 1980. Practical Nonparametric Statistics. 2nd Ed. John Wiley & Sons, New York, NY. 493 pp.

Capital Region District (CRD). 1992. Proposed CRD marine sediment guidelines and quality control/quality assurance guidelines for chemical analyses of marine sediment. Draft rept. prepared by Capital Region District, Victoria, B.C.

Day, R.W. and G.P. Quinn. 1989. Comparisons of treatments after an analysis of variance in ecology. Ecol. Monog. 59: 433-463.

Draper, N.R. and H. Smith. 1981. Applied Regression Analysis. 2nd ed. John Wiley & Sons, New York, NY.

Environment Canada. 1991. Technical guidance manual for aquatic environmental effects monitoring at pulp and paper mills. Volume 2: Methodology (draft report). Prepared by Beak Consultants, Toronto, Ont. 199 pp.

Ferraro, S.P., F.A. Cole, W.A. DeBen and R.C. Swartz. 1989. Power-cost efficiency of eight macrobenthic sampling schemes in Puget Sound, Washington, USA. Can. J. Fish. Aquat. Sci. 46: 2157-2165.

Gauch, H.G. Jr. 1982. Multivariate Analyses in Community Ecology. Cambridge University Press, London. 298 pp.

Gibbons, W.N., M.D. Paine, W.M. Gibson, G.A. Vigers and K. Kingston. 1992. Operational monitoring of the Lesser Slave River. Volume III - 1991. Report prepared for Slave Lake Pulp Corp., Edmonton by EVS Consultants, North Vancouver, B.C. 63 pp. + Tables + Fig. + App.

Green, R.H. 1979. Sampling Design and Statistical Methods for Environmental Biologists. John Wiley and Sons, Toronto, Ont. 257 pp.

Grubbs, F.E. 1969. Procedures for detecting outlying observations in samples. Technometrics 11: 1-22.

Hanson, J.M. and W.C. Leggett. 1982. Empirical prediction of fish biomass and yield. Can. J. Fish. Aquat. Sci. 39: 257-263.

Hoke, R.A., J.P. Giesy and J.R. Adams. 1990. Use of linear orthogonal contrasts in analysis of environmental data. Environ. Toxicol. Chem. 9: 815-819.

Klemm, D.J., P.A. Lewis, F. Fulk and J.M. Lazorchak. 1990. Macroinvertebrate field and laboratory methods for evaluating the biological integrity of surface waters. EPA/600-4-90/030. Environmental Monitoring Systems Laboratory, Office of Research and Development, Cincinnati, OH. 256 pp.

Legendre, P. and M. Fortin. 1989. Spatial pattern and ecological analysis. Vegetatio 80:107-138.

Norris, R.H. and A. Georges. 1993. Analysis and interpretation of benthic macroinvertebrate surveys. pp. 234-286. In: Freshwater Biomonitoring and Benthic Macroinvertebrates. D.M. Rosenberg and V.H. Resh (eds). Chapman and Hall, New York, NY.

Resh, V.H. and E.P. McElravy. 1993. Contemporary quantitative approaches to biomonitoring using benthic macroinvertebrates. pp. 159-194. In: Freshwater Biomonitoring and Benthic Macroinvertebrates. D.M. Rosenberg and V.H. Resh (eds). Chapman and Hall, New York, NY.

Resh, V.H. and J.K. Jackson. 1993. Rapid assessment approaches to biomonitoring using benthic macroinvertebrates. pp. 195-233. In: Freshwater Biomonitoring and Benthic Macroinvertebrates. D.M. Rosenberg and V.H. Resh (eds). Chapman and Hall, New York, NY.

Snedecor, G.W. and W.G. Cochran. 1980. Statistical Methods, 7th ed. Iowa State University Press, Ames, Iowa. 507 pp.

Sokal, R.R. and F.J. Rohlf. 1981. Biometry: The Principles and Practice of Statistics in Biological Research. 2nd Ed. W.H. Freeman and Co., New York, NY. 859 pp.

Tabachnik, B.G. and L.S. Fidell. 1983. Using Multivariate Statistics. Harper & Row, New York, NY.

Wilkinson, L. 1990. SYSTAT: The System For Statistics. SYSTAT, Inc., Evanston, Illinois. 677 pp.

Winer, B.J. Statistical Principles in Experimental Design. 2nd Ed. McGraw-Hill, New York, NY.

Yount, J.D. and G.J. Niemi. 1990. Recovery of lotic communities and ecosystems from disturbance - a narrative review of case studies. Environ. Manage. 14: 547-569.

REPORTING

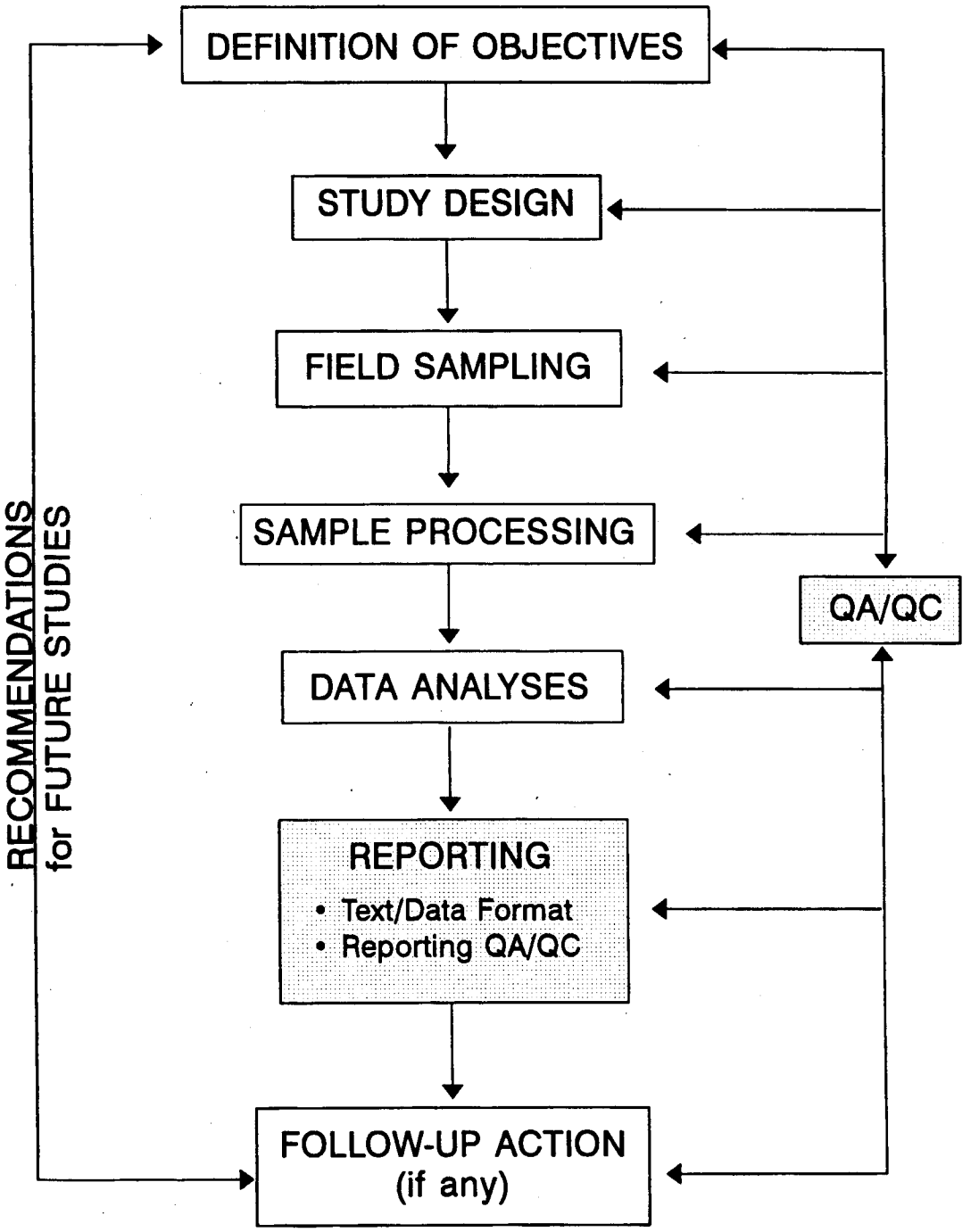

Figure 8. The key elements of a benthos monitoring study. Specific components of reporting have been highlighted.

8.0 REPORTING

8.1 Text and Data Format

Recommendations for reporting were synthesized from suggestions presented at the workshop and from details outlined in the draft EEM guidance manual (Environment Canada, 1991) and Alberta Environment protocol for monitoring benthic invertebrates in rivers (Alberta Environment, 1990).

The report must document all information pertinent to the study and evaluate and interpret the results. Typically, a report consists of the following sections:

1.0 Introduction
1.1 Study Area
2.0 Methods
3.0 Results
4.0 Discussion
5.0 Literature Cited
6.0 Appendices

A Conclusions section and Executive Summary may also be included.

8.1.1 Introduction

The introduction section should include (minimum):

- general background knowledge of the study, i.e., the type of pollution source, brief history of the industry and the current operating conditions, previous reports pertinent to the current study, reasons for conducting a benthos study
- a clear statement of the study objective(s) and testable hypotheses
- brief summary of decision criteria
- statement describing the approach of the study.

8.1.2 Study Area

A full description of the study area must be presented and should include:

- a map of the receiving environment and surrounding area showing the general topography of the area, accurate locations of sampling sites and the source of any pollution discharge, habitat types, locations of other sources or potential interferences (tributaries, other discharge locations, dams, bridges, roads, etc.), legend and scale.

- description of general features of the receiving environment (e.g., for streams - length, discharge, flows; for lakes - area, mean depth)

- description of relevant features of the study area

- general information collected during the reconnaissance survey (physical and chemical characteristics)

- time of sampling.

8.1.3 Methods

This section must describe:

- the study design (including statistical design, preliminary investigations, sampling design, power analysis, sampling frequency and timing and variable selection)

- all field procedures in detail (collecting approach, devices, sieving, labelling, containers, sample preservation, etc.)

- all laboratory procedures (sample handling, sorting, identification, subsampling, reference and voucher collections)

- method of data analysis, and interpretation

- any changes to original protocol or study design (no matter how small)

- Quality Assurance/Quality Control procedures.

8.1.4 Results

This section simply includes the results pertinent to the studies objectives and the hypotheses. Information should be summarized in tables or figures when applicable. All information must be presented without interpretation. Statistical significance must accompany statements of significant differences along with the specific test used and the probability level chosen. The statistical power of tests used should also be provided.

8.1.5 Discussion

The discussion should include:

- an interpretation of the data in relation to the original study objectives and hypotheses
- the type and magnitude of effect
- an evaluation of the data in the context of the original decision criteria
- any recommendations regarding the outcome of the study
- recommendations/modifications to the monitoring program for future studies.

8.1.6 Literature Cited

This section contains an alphabetical list of all literature cited in the text. Formats used in North American journals are suitable.

8.1.7 Appendices

The appendices should include:

- all raw data collected during the study (biological, chemical and physical)
- a complete copy of all field notes
- all QA/QC information (including chain of custody forms)
- detailed results of power analysis calculations.

If the appendices for the study are numerous or lengthy, the appendices can be placed in a separate volume or report prepared to accompany the main study report.

8.2 Reporting Quality Assurance/Quality Control (QA/QC)

The draft report should be reviewed for editorial, grammatical and spelling errors, data entry errors; consistency in format; accuracy of analysis; interpretation of the data; feasibility of recommendations; completeness of QA/QC information and that each section of the report includes the necessary information outlined in the guidance manual. The report should also be reviewed for technical and interpretive content by a senior individual who did not contribute to the study.

8.3 References Cited

Alberta Environment. 1990. Selected methods for the monitoring of benthic invertebrates in Alberta rivers. Water Quality Branch, Pollution Control Division. 41 pp.

Environment Canada. 1991. Technical guidance manual for aquatic environmental effects monitoring at pulp and paper mills. Volume 1: Overview and study design (draft report). Prepared by Beak Consultants, Toronto, Ont. 97 pp.

GLOSSARY

9.0 GLOSSARY

Benthos	In the guideline document, benthos refers to invertebrates living in or on the sediments of aquatic habitats. Other sources will include other organisms (e.g., periphyton) as part of the benthos.
Contrasts	Comparisons of means, usually planned *a priori* (e.g., reference *versus* impact sites). A contrast is a linear combination of the means such that the coefficients assigned to the means sum to 0.
Covariate	An independent variable; a measurement taken on each experimental unit that predicts to some degree the final response to treatment, but which is unrelated to the treatment (e.g., substrate size [covariate] included in the analysis designed to compare benthos abundances upstream and downstream of an outfall).
Data Quality Objectives	pre-defined criteria ensuring that the data are of suitable quality, defensible and meet the studies objectives.
Distance Matrix	A matrix of pairwise distances between pairs of objects (e.g., sites). If measures of similarity rather than distance are used, then the matrix is a similarity matrix. A correlation matrix providing pair-wise correlations among variables is a similarity matrix with the variables as objects.
Effect	Significant and meaningful difference measured in an environmental variable between an exposed and reference area, or operational vs. baseline conditions.
Effluent	Water or complex waste material discharged from facilities into receiving waters. Effluents usually contain chemical contaminants, but may also contain physical contaminants (solids) or have elevated temperatures.
Exposure	AreaAn area of the receiving environment with at least some exposure to effluent or contaminants (e.g., near-field and far-field area).
Far-field Area	An area of contaminant exposure within the zone of influence but at concentrations levels less than the near-field exposure area.
Freshwater Habitat	In the guideline document, freshwater habitats are considered to be lakes, reservoirs, ponds, rivers and streams. The guidelines could be adapted to monitor other freshwater habitats, such as wetlands, marshes, fens, etc.; however, specific characteristics of these habitats (and sampling methods) have not been considered here.
Matrix	Any rectangular row by column array of numbers; a specific data set.
Near-field A	An area of high exposure to the source of potential impact but beyond the region of discharge generated dilution.
Non-Parametric Test	Statistical tests which are independent of the distribution of data; usually based on ranks.

Null Hypothesis	A statistical hypothesis that is a statement of no difference or change. Statistical proof occurs in rejecting but not accepting a null hypothesis.
Parametric Test	Statistical methods that assume the underlying distribution of the data is normal.
Power	The probability of detecting a difference or effect of a given size if it is present.
Quality Assurance (QA)	Refers to externally imposed technical and management practices which ensure the generation of quality and defensible data commensurate with the intended use of the data; a set of operating principles that, if strictly followed, will produce data of known and defensible quality.
Quality Control (QC)	Specific aspect of quality assurance which refers to the internal techniques used to measure and assess data quality and the remedial actions to be taken when data quality objectives are not realized.
Reference Area	An area which is not exposed to a potential source of pollution but exhibits similar natural characteristics to the exposed monitoring sites.
Reference Collection	Collections of specimens used to assist identifications of invertebrates. Investigators should maintain in-house reference collections so identifications can be confirmed and expedited for subsequent studies.
Vector	A line; in the context of this report, a linear combination of variables derived from multivariate analyses of variance (MANOVA) and other procedures. A line on a two-dimensional graph is a vector, because it can be described by a linear combination of *X* and *Y*. Vectors from multivariate analyses are conceptually similar except that they are lines on multi-dimensional graphs.
Voucher Collection	Representative collections of invertebrates, preserved to permit independent verification. Voucher collections are usually deposited with museums, and can be useful to museum researchers for their own projects.

APPENDIX A. USE OF MESOCOSMS

Appendix A. Use of Mesocosms

During the benthos monitoring workshop, there were discussions investigating the suitability of mesocosms for controlled experimental study designs. It was suggested that a monitoring program could not provide all necessary information needed to manage and protect the receiving environment and that mesocosm experiments may successfully augment the monitoring program by investigating specific cause-and-effect relationships. The following is a brief summary of a plenary presentation given at the workshop by Dr. C.J. Perrin (see the workshop proceedings (Gibbons and Booth, 1992)).

The use of mesocosms is gaining wide acceptance for obtaining evidence of responses by aquatic insect communities to chemical manipulations (Allard and Moreau, 1987; Hart and Robinson, 1990; Mundie et al., 1991). When located in the field, they function as open systems and their flow-through design provides for: the exchange of components necessary to represent the stream biogeochemistry, behavioral interactions or organisms, food web interactions, etc. Biological processes are integrated in a trough mesocosm, as they are in a natural stream, thus allowing for considerable extrapolation of findings to larger stream reaches. These systems are quite distinct from "microcosms" that are often used for short term toxicity testing and are defined by Giesy and Odum (1980) as "artificially bounded subsets of naturally occurring environments which are replicable". Mesocosms by definition are larger than microcosms and are continuously colonized with organisms from drift.

Based on the above definition, a mesocosm can be as simple as a series of small flow-through tubes or troughs that can be moved about by a single person. Examples are the chambers used by Peterson et al. (1983) and later adapted by Perrin (1991) for testing effects of chemical manipulations on periphyton growth and biomass. This flow-through design, which incorporates artificial substrata, is also the basis for a more complex experimental trough apparatus known as EXTRA (Bothwell, 1988) that has been used to establish functional relationships between algal growth rates, nutrients, and physical variables for diatom-dominated rivers. Because confounding factors can be controlled or eliminated in a mesocosm, data from one facility can been quantitatively compared with that from another. This approach has been used recently to calculate models of effects of pulp mill effluent and components of the mill effluent on the growth of periphyton in systems that are either N or P limited (Bothwell, 1992).

More complex mesocosms are designed to support insect communities (Mundie et al., 1991; Selby et al., 1985; Perrin, 1991). The flow-through troughs in these facilities are generally larger than those used for periphyton alone and they accommodate a larger range of measurements including benthos abundance, insect drift, and emergence, as well as periphyton growth and biomass and measurement of chemical parameters. The largest mesocosm presently in use is the EPA facility at Monticello, MN (Arthur, 1988) which consists of 8 stream channels that are 520 m in length and include riffle and pool habitats for measurements of abundance, community composition, life history, litter decomposition, as well as basic chemical and physical descriptors. The Monticello facility also supports fish, and there are no restrictions on the type of invertebrates that can colonize the substrata.

With increasing complexity of mesocosm design comes higher cost. Small systems can be kept in a small storage space and used whenever required at virtually any site. Large systems, like the Monticello facility, are obviously fixed and

require a full time staff. Hence, it is important to determine if a small apparatus can meet the demands of the experimental design before spending larger sums of money on a complex facility. In this decision process, it must be kept in mind that there is often a direct relationship between the size of the facility and the complexity of measurements that may be made. However, small facilities such as those used by Mundie et al. (1991) and Perrin (1991) are capable of supporting communities for measurements of abundance, drift, emergence, and other community indices as well as growth and other measures of system function. These parameters cover most short-term requirements. The larger facilities are more amenable to long-term studies which may involve bioaccumulation or those requiring the presence of fish or life history measurements.

It is clear that there are advantages and disadvantages to using mesocosms as a tool within a monitoring program. The greatest advantage is they facilitate the use of experiments to examine causal effects at the community level of system organization, and they assist in answering questions which are commonly left unresolved in a monitoring program. An important distinction between mesocosms and laboratory based experiments is that, because the former are based on community level interactions, they are sensitive to indirect effects of a treatment in addition to direct effects on abundance parameters. An important disadvantage is that the hardware is bulky and, in very large systems, can be expensive. Experiments last a minimum of several weeks but even these are considered short term experiments when examining community interactions. Unless the stream channels are very long, most mesocosms are not designed to examine long term processes such as bioaccumulation or long term (i.e., months or longer) shifts in community processes. Finally, approaches to analyzing data collected from mesocosms are not well established. The general trend is to integrate all organisms in community based analyses rather than concentrate on the mortality or survival of individual taxa.

References Cited

Allard, M. and G. Moreau. 1987. Effects of experimental acidification on a lotic macroinvertebrate community. Hydrobiologia 144: 37-49.

Arthur, J.W. 1988. Application of laboratory-derived criteria to an outdoor stream ecosystem. Intern. J. Environmental Studies 32:97-110.

Bothwell, M.L. 1988. Growth rate responses of lotic periphytic diatoms to experimental phosphorus enrichment: the influence of temperature and light. Can. J. Fish. Aquat. Sci. 45:261-270.

Bothwell, M.L. 1992. Eutrophication of rivers by nutrients in treated kraft pulp mill effluent. Water Poll. Res. J. Canada. In press.

Gibbons, W.N. and A.S. Booth (eds). 1992. Proceedings of the benthos monitoring workshop, February 11-12, 1992, Vancouver, B.C. Prepared for Environment Canada, North Vancouver, B.C. by EVS Consultants, North Vancouver,B.C. 59 pp. + App.

Giesy, J.P.Jr. and E.P. Odum. 1980. Microcosmology: Introductory Comments. In: J.P. Cresy, ed., Microcosms in Ecological Research. Technical Information Center, U.S. Dept. of Energy, Symposium series 52(CONF-781101).

Hart, D.D. and C.T. Robinson. 1990. Resource limitation in a stream community: phosphorus enrichment effects on periphyton and grazers. Ecology 71:1494-1502.

Mundie, J.H., K.S. Simpson and C.J. Perrin. 1991. Responses of stream periphyton and benthic insects to increases in dissolved inorganic phosphorus in a mesocosm. Can. J. Fish. Aquat. Sci. 48: 2061-2072.

Perrin, C. J. 1991. Nutrient limitation of periphyton biomass in a large regulated river of north-central British Columbia. Paper presented at North American Benthological Society Annual Meeting, Santa Fe, New Mexico, May 1991.

Perrin, C. J. 1992. Stream community responses to a gradient of acid mine drainage additions. Prepared for B.C. Acid Mine Drainage Task Force, B.C. Ministry of Energy, Mines and Petroleum Resources, Victoria, B.C. by Limnotek Research and Development Inc., Vancouver, B.C. 44 pp.

Peterson, B.J., J.E. Hobbie, T.L. Corliss and K.Kriet. 1983. A continuous-flow periphyton bioassay: tests of nutrient limitation in a tundra stream. Limnol. Oceanogr. 28:583-591.

Selby, D.A.,J.M. Ihnat, and J.J. Messer. 1985. Effects of subacute cadmium exposure on a hardwater mountain stream microcosm. Water Res. 19:645-655.

Disclaimer

This publication contains the results of a project conducted under contract to Environment Canada. The ideas and opinions expressed herein do not necessarily state or reflect those of Environment Canada.