Guidelines for Designing and Implementing Aquatic Effects Monitoring Programs for Development Projects in the Northwest Territories

Recommended Procedures for Developing Detailed Designs for Aquatic Effects Monitoring Programs

AEMP Technical Guidance Document Volume 4

Indian and Northern Affairs Canada
Yellowknife, Northwest Territories

June 2009 Version

Guidelines for Designing and Implementing Aquatic Effects Monitoring Programs for Development Projects in the Northwest Territories

Recommended Procedures for Developing Detailed Designs for Aquatic Effects Monitoring Programs

AEMP Technical Guidance Document
Volume 4

June 2009 Version

Prepared by:

B. Zajdlik[1], D.D. MacDonald[1], and INAC Water Resources[3]

[1]Zajdlik & Associates Inc.
R.R. # 3
Rockwood, Ontario N0B 2K0

[2]MacDonald Environmental Sciences Ltd.
#24 - 4800 Island Highway North
Nanaimo, British Columbia V9T 1W6

[3]Water Resources Division
Indian and Northern Affairs Canada
P.O. Box 1500
Yellowknife, Northwest Territories X1A 2R3

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Acronyms

AEMP	-	Aquatic Effects Monitoring Program
BACI	-	before and after control impact
BACIP	-	before and after control-impact paired designs
CCME	-	Canadian Council of Ministers of the Environment
CEAA	-	Canadian Environmental Assessment Act
CI	-	control impact
DQO	-	data quality objective
EEM	-	Environmental Effects Monitoring
EQG	-	environmental quality guideline
EQO	-	environmental quality objective
FSP	-	field sampling plan
GIS	-	geographic information system
GLWB	-	Gwich'in Land and Water Board
HSP	-	health and safety plan
INAC	-	Indian and Northern Affairs Canada
K_{oc}	-	organic carbon partition coefficient
K_{ow}	-	octanol water partition coefficient
LWB	-	the Land and Water Board
MRP	-	Management Response Plan
MVEIRB	-	Mackenzie Valley Environmental Impact Review Board
MVLWB	-	Mackenzie Valley Land and Water Board
MVRMA	-	Mackenzie Valley Resource Management Act
NWTWA	-	Northwest Territories Water Act
NWTWB	-	Northwest Territories Water Board
NWT	-	Northwest Territories
QAPP	-	quality assurance project plan
QA/QC	-	quality assurance/quality control
SLWB	-	Sahtu Land and Water Board
TK	-	Traditional Knowledge
USEPA	-	U.S. Environmental Protection Agency
VEC	-	valued ecosystem component
WLWB	-	We'eezhii Land and Water Board
WQG	-	water quality guideline
WQO	-	water quality objective

1.0 Introduction

Considerable effort has been expended in Canada and elsewhere worldwide to develop guidance for monitoring the effects of human activities on aquatic ecosystems. These efforts have resulted in a variety of guidance documents that could be used to support the design of Aquatic Effects Monitoring Programs (AEMPs; e.g., Ecological Monitoring and Assessment Network, Environmental Effects Monitoring, International Organization for Standardization, United States Environmental Protection Agency). Project proponents are encouraged to review such guidance documents during the AEMP design process, but not all aspects of these documents can be applied directly to the development of AEMPs within the NWT. The sensitive nature of the northern environment and its ecosystems, coupled with traditional use of aquatic resources in the region influences the applicability of these guidance documents to the Northwest Territories.

Selecting an appropriate AEMP design is a critical step in designing a defensible AEMP but it is only one of several other equally critical steps in the AEMP design process. For example, unless a problem is well defined in the context of the receiving environment and environmental stressors, and expressed in terms of the concerns of Aboriginal governments/organizations, federal and territorial governments, regulatory boards and other interested parties, a technically well designed AEMP may produce irrelevant information. Therefore, it is necessary that this design step be used in conjunction with other steps to design an AEMP that avoids collection of information that does not contribute directly to meeting the objectives of the monitoring program. Similarly, the involvement of the AEMP Working Group (see the AEMP Guidelines Overview Report for more information on the recommended use of an AEMP Working Group) during detailed AEMP design development is recommended.

This Technical Guidance Volume is intended to provide guidance on the detailed design of AEMPs in the Northwest Territories (NWT). The steps in the detailed AEMP design process include:

1. Select the appropriate monitoring program design;

2. Select sampling locations;
3. Choose appropriate effects sizes;
4. Determine the necessary sample size; and,
5. Select appropriate sampling frequency.

Following a description of each of these five steps, additional considerations for designing detailed monitoring designs are discussed, including selection of reference stations, evaluation of variability, and application of statistical analyses. Finally, AEMP adaptations are briefly discussed.

2.0 Selecting a Monitoring Program Design

The first step in selecting an AEMP design is understanding the differences between specific monitoring designs that have been used in monitoring aquatic environments in Canada and elsewhere worldwide. Various monitoring designs have been developed to guide sampling in uncontrollable environments. All of the monitoring designs result in the collection of data in the exposure area and compare it with data representing no exposure to site-related stressors. The data representing no exposure to stressors of potential concern may correspond to the exposure area before exposure began (i.e. baseline information that represents a temporal reference) and/or to a different geographic area (a spatial reference). The monitoring designs discussed within this section were developed to evaluate the effects of point source discharges; however, with some adaptation, the concepts are applicable for evaluating effects associated with non point discharges. Note that guidance on the selection of groundwater sampling locations is beyond the current scope of this document. The following monitoring designs are discussed below:

- Control-Impact Design;
- Before After Control-Impact Design;
- Before and After Control-Impact Paired Design;

- Asymmetrical Before and After Control-Impact Design;
- Gradient Design;
- Regional Reference Design; and,
- Alternate Monitoring Strategies.

Although these monitoring designs are based on western science, incorporation of Traditional Knowledge (TK) can be very valuable when selecting a monitoring program design. TK can help determine appropriate monitoring locations and timing in both reference and exposure areas, while ensuring the most appropriate valued ecosystem components (VECs) are monitored.

2.1 Control-Impact Designs

Control impact (CI) designs are the simplest AEMP designs. This approach to monitoring receiving water systems is indirectly recommended by the Environmental Effects Monitoring (EEM) program (Environment Canada 2002; 2004), which states: "The reference area should be located in the same water body as the effluent discharge". The exposure area is defined as "all fish habitat and waters frequented by fish that are exposed to (mill) effluent for surface water" (Environment Canada 2004). "Near-field" and "far-field" exposure areas are found within the general exposure area.

In the CI design, data from the exposure area are compared to data collected within the reference area. Effects on the aquatic environment associated with the project are identified when statistically significant differences in one or more measurement endpoints are observed between the two areas. Reference and exposure areas are chosen following the criteria discussed Section 7.1 and Appendix 1, respectively.

Sites within the exposure area represent a continuum from the highest effluent concentrations immediately outside the initial mixing zone to virtually undetectable concentrations at the far field. Thus, classification of stations within the "near field" and "far field" areas (and in some instances mid field areas if applicable) can result in considerable variation in exposure. To the extent reasonable, variation in exposure concentration (variability) within exposure sub categories such as "near field" and "far field" should be minimized to increase statistical power (discussed in Section 7.3) without increasing sample sizes (i.e., by positioning sampling stations within an area at similar depths and at similar distances from the outfall and by collecting samples at the same time. Inclusion of mid field stations may assist in controlling variations in exposure. The number of samples to collect within each area or area sub category is discussed in Section 2.6 of Technical Guidance Document Volume 3 (Specifying Performance or Acceptance Criteria) and Section 5.0. Also, see Gorman and Karr (1978) Vannote *et al.* (1980), Voelz and McArthur (2000), and Flotemersch *et al.* (2006) for more information.

The primary advantage of the CI design is that it provides a temporally-relevant basis for evaluating project-related effects (i.e., exposure and reference areas are sampled simultaneously, thereby minimizing the probability that observed differences reflect temporal variability). Accordingly, this design can have high power for detecting the effects of a project. One of the main challenges of this design is identifying areas that have similar physical, chemical, and biological conditions relative to the exposure areas (near-field, mid-field, and far-field).

2.2 Before and After Control-Impact Designs

Before and after control impact (BACI) is a modification to the basic CI design. BACI follows the CI design by collecting samples in a reference area(s) and an exposure area(s), but extends the design by also collecting samples from the same areas (reference and exposure) before the effluent discharge occurs. Environmental samples are collected simultaneously at exposure and reference sites at fixed frequencies. The intent of a BACI designs is to ensure that changes in a measurement endpoint within an exposure area are due to the exposure and not natural changes in the environment over time (see Green 1979 for statistical analysis method; Underwood 1991; 1994).

In the NWT, relatively little baseline or pre-development information is available for most watersheds. Therefore, proponents must collect baseline data prior to project development. To be useful for BACI evaluations of project-related effects, baseline sampling programs must focus on the same measurement endpoints that are targeted in the AEMP. This can be achieved by initiating problem formulation early in the project development process and utilizing the preliminary problem formulation to design the baseline monitoring program. See Technical Guidance Document Volume 2 for more information on problem formulation.

The primary advantage of BACI designs over CI designs is that general shifts in environmental conditions can be attributed to a stressor. A disadvantage of this method is that of pseudoreplication, as is the case with CI designs. The following sections describe modifications to the basic BACI design that attempts to address this issue. Smith (2002) provides an excellent description of BACI design, including relevant information on the analysis of the resultant data.

2.2.1 Before and After Control-Impact Paired Designs

Before and After Control-Impact Paired Designs (BACIP) are intended to address the limitations of the BACI design. Using a BACIP design, samples are collected in the reference and exposure areas. The difference between the two measurement endpoints collected at any given time becomes one observation. This procedure is repeated at several randomly selected times (so as to avoid the possibility of samples coinciding with cyclical patterns). In this way, observed differences between the exposure and reference sites can be more confidently described as project-related effects. This procedure is thought to address the criticism (Hurlbert 1984) of non-random sampling in the BACI design (Stewart-Oaten et al. 1986).

2.2.2 Asymmetrical Before and After Control-Impact Design

Asymmetrical BACI designs are described by Environment Canada (2002) as a “multiple control-impact” study. This design is advocated when a variable that can affect the response being measured (i.e., a confounding variable) in a reference area varies from that for the exposure area. Underwood (1991) suggested increasing the number of reference sites to deal with random oscillations of the reference condition. These multiple reference sites may be collected under a CI or a BACI design.

If Asymmetrical BACI designs are used, care must be taken when selecting the multiple reference areas such that variability among reference areas truly reflects variability at the local scale. Variability measured at the non-local scale can obscure changes at the local scale (Stewart-Oaten and Bence 2001).

2.3 Gradient Designs

In the event that rapid dilution prevents the selection of sites within subcategories of the exposure area (e.g., “near-field”) that are relatively homogenous with respect to exposure (or if there are no suitable reference areas), a gradient design may be used. In this design, samples are not collected within sub-categories of the exposure area but rather at increasing distances from the mixing zone. In this context, distance does not necessarily refer to geographical distance, but rather to distance along the path the effluent travels. When there is a strong primary concentration gradient due to currents or channels (such as in rivers), samples should be collected along the primary axis. If effluent is released into a large waterbody, such that there is a radial concentration gradient, samples should be collected along transects or “spokes” with the central point being the effluent source. This is known as a radial gradient design.

Where the gradient is a simple gradient or radial gradient, sampling sites should be selected such that the entire concentration gradient is spanned. The optimal allocation of stations is not likely equispaced along the concentration gradient but instead with a majority of sample locations concentrated in areas of higher effluent exposure.

In a CI design, "effects" due to the effluent are ascribed due to a statistical difference between reference and exposure areas. In gradient designs, "effects" due to the effluent are ascribed to a non-random change in measurement endpoint with increasing distance from the mixing zone. This is equivalent to a non-random change in measurement endpoint due to reduced exposure since exposure decreases with "distance".

One potential challenge of sampling in a receiving environment that dictates a gradient design is the likelihood that distance from the source (regardless of gradient design sub-type) is correlated with other environmental variables. In a stream or river (lotic) environment this might be the channel unit (see Appendix 1 for more details). In a lake (lentic) environment, water depth is often a confounding factor. To the extent possible, such confounding effects should be avoided.

If variables that confound the exposure cannot be avoided (e.g., for example, discharge occurs in a shallow area of a lake and a denser effluent plume following contours increases with depth inducing a depth gradient in the exposure area), a reference gradient with the same confounding variable may be used as a basis for comparison. Monitoring designs of this type are known as multiple gradient designs. When confounding factors cannot be separated from the exposure gradient using a multiple gradient design, a statistician should be consulted.

2.4 Regional Reference Designs

The regional reference approach (Bailey *et al.* 2003) is fundamentally different than other approaches in that a predictive model is developed and used as a basis for comparison. In the regional reference design, biotic assemblages are collected from sampling stations designated as reference sites and thought to define the "reference condition". Then, using multivariate statistical tools, groupings of organisms associated with non biological variables are identified. A biologic assemblage at an exposure site is predicted on the basis of the non biological variables identified as being important drivers of biotic assemblages (in the reference condition). If the biotic assemblage at an exposure site differs from that expected, the difference is ascribed to the exposure. The Canadian Aquatic Biomonitoring Network (CABIN) represents an example of a regional reference monitoring design (Reynoldson *et al*. 2003).

2.5 Alternate Monitoring Strategies

In rare cases, none of the designs discussed above may be appropriate for a site. The designs might not be implementable due to safety reasons, an inability to separate effluent effects from other effects (although this issue will rarely arise in the North where site facilities are very widely spaced), or other reasons. Environment Canada (2002; 2004) discusses alternative monitoring strategies (e.g., these include the use of caged mussel and mesocosms studies).

Despite the change in monitoring strategy, the experimental designs themselves can be the same as those used for field sampling. For example, a mesocosm study may use reference water and effluent to create a control impact experimental design. The effluent might be serially diluted to create a gradient experimental design. Regardless of the physical sampling program, the confounding variable concept still applies, as will other concepts such as Type I and II error rates, effect sizes, etc. discussed later in this document.

2.6 Summary of Monitoring Program Designs

The monitoring designs presented above have various advantages and disadvantages which are summarized in Table 1. An important component of many of the above monitoring designs is the establishment of a "non-exposure" or reference site(s). A physical reference area may be in the same or another waterbody as the exposure area. In some special cases, where a suitable reference area is not available, data may be collected only from an area with an exposure gradient. What makes a reference area "unsuitable" are factors that may obscure or modify the data collected (i.e., physical and or chemical characteristics). These factors and their importance in designing aquatic effects monitoring programs are discussed further in Section 7.0.

The control impact monitoring design is recommended for use when appropriate due to its simple conceptual design and low relative cost. Although the conceptual design is simple, it is not simplistic (i.e., the design does not sacrifice scientific rigour to achieve simplicity).

The simple gradient approach is also conceptually straightforward but, given variability in many biological measurement endpoints, it may be difficult to detect effects if sample sizes are small. Also, if the effluent comprises a large proportion of the water moving through the receiving water body, it may become very difficult to avoid confounding variables. More complicated gradient designs can address these issues but can be increasingly difficult to interpret.

When reference sites are unavailable or the exposure gradient is badly confounded, alternative monitoring programs might be necessary. Note that these alternative programs may still fall into a general control-impact type of design because "exposure" data are compared to "reference" data. A flowchart summarizing the monitoring design selection process is presented in Figure 1.

3.0 Selecting Appropriate Sampling Locations

The second step, selecting sampling locations, represents a key element of the overall AEMP design process. The selection of sampling locations needs to occur on two scales (large and small) to ensure that valid comparisons can be made across groups of locations. This section describes the factors that need to be considered during the sampling location selection process. The discussion within this section uses terminology (exposure area, reference area, near field, mid-field and far field) described above in Section 2.1. The discussion assumes a point source discharge via an effluent or generically, a "stressor". With some adaptation, the concepts are applicable to some non point discharges.

3.1 Large Scale Selection of Sample Locations

The first level of large-scale aquatic sample selection is at the level of a terrestrial construct the ecoregion. This is because an ecoregion is a set distinctive regional ecological factors, including climate, physiography, vegetation, soil, water, and fauna that, acting together, affect biological communities within the ecoregion in a similar manner (Marshall and Schut 1999). Corkum (1992) and Hughes *et al.* (1994) found that aquatic biotic assemblages vary more among ecoregions than within ecoregions. For this reason, Environment Canada (2004) and USEPA (1998) recommend that exposure and control samples for aquatic programs be selected from the same ecoregion.

Ecozones exist within ecoregions and are defined by Marshall and Schut (1999) as "A subdivision of an ecoregion characterized by distinctive assemblages of relief, landforms, geology, soil, vegetation, water bodies and fauna". The selection of sampling locations within the same ecozones will almost certainly reduce the effects of large scale confounding factors. A detailed graphical presentation of the ecozones and ecoregions of the NWT and Nunavut is available at: http://sis.agr.gc.ca/cansis/nsdb/ecostrat/printed_maps.html.

From the hydrologic perspective, large scale selection of sampling locations should occur within the same watershed (or drainage basin). A watershed is an area of land defined by the common movement of the collected precipitation and groundwater in a downstream (downhill) direction to a single waterbody. In the NWT, there is one primary watershed, the Mackenzie River Basin. However, there are two primary water bodies that surface waters in the NWT feed, Hudson Bay and the Arctic Ocean. Each of the basins within the NWT may be broken down into smaller sub basins. AEMP sample locations should be selected within the smallest level of sub basin that is practical.

3.2 Small Scale Selection of Sample Locations

The selection of sample locations at the small scale is the final opportunity to control potentially confounding variables and is critical for a defensible AEMP. This site selection is of such importance that different criteria exist for water that is flowing (lotic), standing (lentic), coastal, estuarine, and even for sub divisions of these categories. Criteria for small scale site selection also vary by the type of VEC as the relative importance of some confounding factors can vary (such as substrate type for benthic macroinvertebrate communities). Appendix 1 synthesises the recommendations of existing guidance for small scale selection of sample locations. Personnel developing AEMPs should consider not only the criteria presented therein, but other criteria that may be relevant at the site or on a project specific basis.

One criterion that warrants special attention is that of stream reach. Stream reaches are defined by expert opinion and definitions are regionally specific. Therefore, a set of criteria most relevant to the NWT should be defined by aquatic biologists, geologists, hydrologists, hydrogeologists and interested parties. This is discussed in the section entitled "Reaches" in Appendix 1.

4.0 Choosing Appropriate Effects Sizes

The third step in developing an appropriate AEMP design is choosing the most appropriate effect sizes. Effect sizes are the changes in a measurement endpoint that are unacceptable on the basis of any combination of scientific, social or political values. An example of a scientific valuation is a change of more than two standard deviations in a metric, such as benthic macroinvertebrate taxonomic richness, for a reference area. An example of a sociologic valuation is "no effects" on fishing. Each of these valuations can be converted into a testable hypothesis that can be evaluated using the results of a well-designed AEMP. For example, the sociologic valuation of "no effects" on fishing might become the following testable null hypothesis:

H0: There is no difference in catch per unit effort between *the exposure and reference areas.*

In this example, the effect size is 0 (zero), i.e. there is no acceptable difference in catch per unit effort.

Environment Canada (2004) discusses three methods for generating effect sizes (in order of preference):

1. Set an effect size based on *a priori* knowledge of what percent change in an ecosystem causes long-term negative changes in the environment.
2. Using data from a reference area(s) within the same ecoregion to set an a priori effect size.
3. Set an effect size at ± 2 standard deviations using data from a reference area(s).

Of these three methods, Environment Canada (2008) used the third option to provide a set of recommended effect sizes (Table 2) for use when assessing potential environmental effects from metal mine or pulp and paper mill effluents in the context of the EEM programs. The effect sizes in Table 2 may serve as a starting point for discussions on acceptable effect sizes that occur during AEMP development.

Table 2 is based upon the performance data of pulp and paper mills of varying ages and with varying levels of effluent treatment using the second cycle of pulp and paper EEM program data. The effect sizes were chosen considering "statistically significant effects in at least one of the core measurements" (Environment Canada 2008). Thus, the selection of effect sizes was driven by statistically detectable differences, given the sampling requirements in place at that time using data from pulp and paper mills with varying levels of treatment. These effect sizes do not reflect the method recommended by Environment Canada (2004); namely effect sizes that correspond with unacceptable ecological changes. Also, the critical effect sizes presented in Table 2 were chosen by Environment Canada (2008) "in the absence of clear scientific understanding of the long-term implications of these effects".

It is important to note that in the NWT, careful consideration should be given to the size of ecological change that is of long term concern. In the absence of the knowledge required to quantify such a change, a precautionary approach should be taken. Aboriginal governments/organizations, regulatory boards, federal and territorial governments and interested parties should also understand that the choice of effect sizes (along with the Type II error rate - Section 7.3) influences the cost of an AEMP. Mapstone (1995) suggests that the choice of effect size should be a primary driver of a monitoring program design.

4.1 Detectable Effect Sizes

Once data have been collected following implementation of an AEMP, the question may arise what is the smallest difference or “effect size” that the program could detect? This is variously described as the “minimum detectable difference” or “minimum significant difference”. This difference is the smallest difference in a VEC that would be labelled as “statistically significant”, given the Type I error chosen *a priori*. For example, consider a Control-Impact design where an effect size was set as ± 2 standard deviations (of the controls site) for a water quality variable such as nickel. If the design was actually able to detect as small a difference as ± 1.6 standard deviations (between control and exposure locations), then the design is more powerful than originally intended. The converse is also true; if the design cannot detect a difference as small as ± 2 standard deviations, the monitoring program is insufficiently powerful to detect changes of expressed interest.

The minimum detectable, or minimum significant, difference is a useful quantity to estimate when assessing how meaningful AEMP conclusions are. Methods for estimating the minimum detectable difference may be found in general statistical textbooks such as Sokal and Rolf (1995), Steel *et al.* (1997), and Zar (1999).

5.0 Determining the Necessary Sample Size

The fourth step in developing an appropriate AEMP design is determining the required sample size. Sample sizes are driven by the required statistical power (Section 7.3) of the monitoring program and the recommended effect sizes (Section 4.0). Since environmental data are inherently variable, the number of samples needed to evaluate effects of specific size can only be estimated. The equations used to estimate sample sizes are directly related to the statistical significance test conducted. For example, if a control impact type hypothesis is tested using a Student T test, the formula used to generate the test statistic is inverted. Then, using the appropriate Type I and II error rates, the required sample size is estimated. In order for these sample size estimates to be valid, the assumptions of the statistical test used must also be valid. The assumptions for the Student T test are that the variances between the reference and exposure group measurement endpoints are equal (unlikely for environmental data when there is a substantive ecological effect) and that data are normally (i.e., Gaussian) distributed (again, an unlikely distribution for many commonly encountered measurement endpoints). See Section 7.3 for further discussion on statistical considerations.

The sample sizes needed to achieve a given statistical power and a given effect size will vary among measurement endpoints. For example, Angermeier and Smogor (1995) found that greater sampling effort is necessary to characterize fish species richness than abundance. Accordingly, different levels of sampling effort among measurement endpoints may be required.

Sample size formulae for the most commonly encountered statistical hypotheses are presented in general statistical textbooks such as Sokal and Rolf (1995), Steel et al. (1997), and Zar (1999). The sample sizes consistent with effect sizes of Environment Canada (2008; Table 2) are presented in Table 3 for various combinations of Type I and II error rates. However, the discussions regarding the choice of effect sizes (Section 4.0), Type I and II error rates (Section 7.3.1), and statistical assumptions should be kept in mind before adopting the sample sizes presented in Table 3.

6.0 Selecting the Appropriate Sampling Frequency

The fifth step in developing an appropriate AEMP design is selecting the appropriate sampling frequency. This step is extremely important and is based in part on the culmination of the previous steps (Step 1 through Step 4). In addition, sampling frequency will also be dependant on the variability of natural conditions in the area.

It is known that any receiving environment will change with season and the nature of change is a function of geomorphology (both surface area to volume ratios and basin morphometry affect stratification and, hence, turnover), hydrology (retention times change as flow conditions change), ice cover (limits mixing and oxygenation), prevailing winds (affects mixing, aerial deposition, long range transport), and other factors. The influence of other developments such as dams, roads and or sewage outfalls on the physical and/or chemical characteristics of a receiving water body can also change temporally. Additionally, industrial processes themselves may change seasonally as a consequence of seasonal restrictions on various site activities.

An AEMP is intended to monitor possible effects of stressors on the environment. While continuous monitoring of all of the physical, chemical, and biological characteristics of the aquatic ecosystem would provide comprehensive data for evaluating project related effects, it is not possible to implement such a monitoring program. For this reason, many monitoring programs are designed to evaluate worst-case conditions, when exposure and/or biological responses are greatest. Some of the factors that are considered in identifying worst-case conditions include:

- Levels of abiotic, non-process related stressors (e.g. low dissolved oxygen);
- Levels of process-related stressors (e.g., effluent concentrations); and,
- Periods of biological sensitivity (e.g. swim up stages for fish) or maximum diversity.

Due to the large number of factors, it is unlikely that the worst case conditions can be adequately defined in advance of actual monitoring. Therefore, monitoring programs should be designed to evaluate seasonal and spatial variability in stressors and responses. Such information, collected over a three to five year period, can be used to focus the monitoring program in subsequent years. Therefore, sampling frequency should generally be high during the first three to five years of monitoring, and reduced thereafter if it is determined that data collected at certain times of the year are redundant.

7.0 Considerations for Developing Detailed Monitoring Program Designs

There is a series of other considerations that must also be worked into the decision-making process when designing an AEMP. Many of these items are directly linked to the five steps described in Section 2.0 through Section 6.0. These considerations are important for ensuring that the monitoring program is sufficiently robust and effective, but also not overly cumbersome. The following section will outline these considerations to assist practitioners in designing effective and efficient AEMP. These considerations include:

- Determining appropriate reference stations and conditions;
- Addressing variability, including that associated with natural process and with developmental activities;
- Statistical, including statistical power and determination of acceptable error rates; and,
- Other monitoring considerations, including cooperation, synoptic sampling, and historical data use.

7.1 Determining Reference Stations and Conditions

Measurements taken at reference stations are the basis from which changes in exposure conditions are measured (for non gradient type monitoring designs). Inappropriate selection of a reference station or stations will result in biassed and/or incorrect conclusions. Environment Canada (2004) suggests that reference stations represent areas "in which impacts are lowest or disturbance is minimal". In the NWT, where most lakes are subject only to contamination from long range transport and changes due to other activities or uses are minimal, reference lakes or streams will generally be relatively pristine.

At the broadest scale, reference areas must be selected in the context of the physical receiving environment. If the receiving environment is a river, an upstream reference area may be appropriate. In a lake environment, an isolated embayment or a separate water body might be selected as a reference area. The ideal reference area is one that is similar in all respects to the area receiving effluent, but with the absence of effluent. In a controllable experiment, the receiving water body would be replicated without effluent release. Since no water body is perfectly replicable, criteria must be applied to ensure unbiased conclusions when selecting reference site and/or reference areas. Such criteria are specific to the physical, chemical, and biological nature of the receiving environment.

7.1.1 Scales of Reference Station Selection

Reference stations may be selected at a local or regional scale. The selection of reference stations at a local scale is preferred to minimize the possible effects of confounding factors. Environment Canada (2004) makes this statement in another way: "The reference area should be located in the same water body as the effluent discharge". If reference stations at a local scale are not available, then there are several options for selecting bases for comparison:

- Select reference stations that are within the same ecoregion and in comparable drainage basins (Hughes *et al.* 1994). This is the regional reference approach;
- Create a theoretical reference condition (applies only to benthic macroinvertebrates; Flotemersch *et al.* 2006); or,
- Select exposure stations so that there is a systematic decrease in effluent exposure (i.e., the gradient approach). In this case, comparison is not strictly with a reference condition but rather an implied reference condition defined by extrapolating the relationship between a measurement endpoint and some variable associated with the effluent such as aqueous chemical concentration, sediment chemical concentration, distance from source, and others.

7.1.2 General Reference Stations

This section discusses general selection criteria for AEMP reference stations. Criteria specific to the lotic (river) and lentic (lake) receiving environments are discussed Appendix 1, Sections A1.2 and A1.3, respectively. USEPA (1998) and Environment Canada (2004) suggest a hierarchical terrestrial classification to create the context in which reference areas are selected. The Environment Canada (2004) scheme (modified from Maxwell et al. 1995) begins with identification of the ecoregion in which the exposure area is found. The large-scale selection of sampling areas (ecoregion or ecozone > drainage basin or watershed) was discussed in Section 3.1. Reference stations should be selected from the same drainage basin as exposure stations if possible. Otherwise, reference stations should be selected from the "nearest comparable drainage with minimal development" (Environment Canada 2004).

At a smaller scale of reference area selection, the surrounding land use and/or surrounding vegetation should be considered. In the NWT, there has not been much development and therefore, not much land use that would change biotic assemblages.

However events such as fire (whether natural or not) does cause changes in water chemistry (Kokelj *et al*. 2009) and aquatic biota (Cott *et al.* In press).

In some special cases, where a suitable reference area is not available, data may be collected from an area with an exposure gradient. What makes a reference area "unsuitable" are factors that may obscure or modify the data collected (i.e., physical and or chemical characteristics). These factors and their importance in designing aquatic effects monitoring programs were discussed in Section 3.0.

7.1.3 Riverine Reference Stations

General criteria for selecting reference stations in rivers are discussed in Appendix 1. Selecting reference areas within a riverine environment is dependent on a number of factors, such as; stream order, habitat and substrate type, biological diversity, and hydrological characteristics. The most relevant factors or criteria for establishing appropriate reference stations within a northern river should be defined by experts (i.e., TK holders, aquatic biologist, hydrologists).

7.1.4 Lake Reference Stations

In addition to the general criteria described in Appendix 1, USEPA (1998) discusses five criteria to select suitable lake reference stations. These are: 1) expert consensus; 2) biological survey of sites; 3) paleolimnology; 4) use of historical data; and, 5) prediction of expected conditions using ecological models. Of these criteria, the latter three may be less useful in the NWT. Sufficient historical data within a given waterbody are generally not available and validated ecological models or the data needed to calibrate them are not likely available.

Expert consensus in the NWT should include physical scientists, aquatic biologists, resource managers, fisheries biologists and traditional knowledge holders. When there are few lakes of a similar size or class (such as Lac de Gras) in the NWT, USEPA (1998) suggests that a number of "most similar" lakes be studied and the best conditions for each variable from the set of lakes be used to define the reference condition.

7.2 Addressing Variability

The VECs measured within an AEMP are all subject to variability. Water quality conditions are known to vary substantially on temporal and spatial bases. Biotic indices change both intra annually and inter annually. Sediment quality is highly variable in the micro scale, particularly when high molecular weight organic compounds are being measured.

This underlying variability in VECs makes it challenging to distinguish patterns from the apparent noise in the system. For this reason, it is important to incorporate monitoring program design elements that facilitate characterization of the various sources of variability and support interpretation of the resultant data. Some general sources of variability and strategies to address them are identified below; however, this list is not exhaustive and should be augmented, particularly with site specific knowledge.

- ***Small-scale spatial variability*** - Collect and composite multiple samples (i.e., sub-samples) to smooth out micro scale heterogeneity which is of little interest. This applies to biotic indices, water and sediment quality [Environment Canada (2004) discusses the distance between samples such that the samples may be designated as "replicates" rather than sub-samples].
- ***Variability with depth*** - Conduct grab sampling at various depths and compare the results using statistical methods.
- ***Cross-sectional variability*** - Conduct a station evaluation to determine if water quality varies significantly along a cross section of a river. Data collected along the transect are compared using statistical methods.

- ***Seasonal variability*** - Conduct sampling over the course of a year, with intensive sampling conducted during freshet events and the open-water period. Use graphical and statistical techniques to evaluate seasonal variability.
- ***Interannual variability*** - Conduct long term sampling at a fixed frequency at fixed stations. Use graphical and statistical techniques to evaluate long term trends.
- ***Variability due to regional and global influences*** - Conduct long term monitoring at a fixed frequency at fixed reference stations. Use graphical and statistical techniques to evaluate long term trends.

Acquisition and evaluation of data on the sources of variability in VECs provide a means of designing and optimizing AEMPs.

7.3 Statistical Considerations

Statistical tools may be used to test hypotheses within an AEMP. Statistical tools are used to interpret AEMP data because they are objective, enable definitive statements to be made despite variability in data, and can be used to determine how certain we are regarding a statement. Hypotheses are statements that express a belief about a statement. This belief can be stated in a direct manner, such as “the project under consideration is causing no harm to the environment”. The hypothesis or belief is tested by collecting evidence or data. After examining the data collected, the initial belief might be rejected and an alternate conclusion may be reached (e.g., “the project is having a harmful effect on the environment”). Alternatively, after examining the data the initial belief might be accepted and the conclusion reached: there is “no effect”. Some examples of translating statements of belief into testable hypotheses are presented below prior to a more formal treatment of hypothesis testing.

Example 1: Using Traditional Knowledge

A long history with the land and water enables TK holders to make statements such as: "We used to catch a lot of lake trout there" (pointing to a location on a map). This statement of traditional knowledge can be translated into a testable hypothesis as follows:

Use specific fishing methods for a specific amount of time at the site of concern and in a similar lake or location that is not exposed to the site (effluent, dust, etc.) and count the number of fish caught. If the number of fish caught near the site is much lower than the number from a different lake or stream there may be a problem. The testable hypothesis for this example might be:

There is no difference in the amount of fish caught at the two locations.

The alternative hypothesis is:

There is a reduction in fish caught at the site of concern compared to the reference location.

Once a difference in the catch rates in fish is noted management actions would be necessary. Such management actions would be described in the management response plan (MRP) that is developed for the project.

Example 2: Human Health

This example addresses the concern often expressed in the NWT regarding the safety of drinking water. A relevant testable hypothesis might be:

The concentration of copper in the Yellowknife River does not exceed the Canadian water quality guideline (WQG) for the protection of human health.

Here the testable hypothesis uses the WQG as a basis for comparison. The WQG, when used in this manner, represents the Action Level. The alternative hypothesis is:

The concentration of copper in the Yellowknife River exceeds the Canadian WQG.

In this example, exceedance of the Canadian WQG would likely trigger a management response (if the WQG was selected as an Action Level in the Management Response Plan (MRP; the MRP is the new term that will be used to replace the Adaptive Management Plan. This new term is also being used by the land and water boards). Action Levels should be set so as to allow sufficient time for management response plans to take effect prior to occurrence of a significant deleterious effect. Action Levels are discussed in Technical Guidance Document Volume 3.

The examples presented above present specific testable hypotheses. These examples can be generalized in preparation for a more detailed discussion of testable hypotheses and AEMP design. A general null hypothesis (or statement of belief, designated as H0) relevant to AEMPs is:

There is no effect of the effluent (stressor) on a specific measurement endpoint.

Given enough evidence to the contrary, this hypothesis or belief may be rejected in favour of an alternative hypothesis (designated as Ha):

There is an effect of the effluent (stressor) on a specific measurement endpoint.

Statistical tools are however “blind” in that, conclusions are only as good as the data used to make inferences. Therefore every aspect of AEMP design leading up to a conclusion merits the same attention that the conclusion will receive.

7.3.1 Making an Incorrect Conclusion

A discussion illustrating how an incorrect conclusion can be reached using data collected from a variable environment is presented below. Consider a single measurement endpoint collected at a specific time within an AEMP, such as the concentration of a chemical. Its concentration will vary from sample to sample, randomly around some central concentration (e.g., mean, median). The set of chemical concentrations can be variously summarized; for example the median might be used to represent the "middle" or "centre" of the dataset. If another set of data were collected at slightly different locations or on a different day, a slightly different median would be produced. Repeated sampling will produce a variety of medians clustered around some central value; therefore, the value obtained from a specific collection of data is only one of the possible sets.

The conclusion reached following a hypothesis test is a function of the particular median values estimated from the reference and exposure areas (in a CI design). It is possible that even if an effluent produces an elevated chemical concentration, a very low median concentration might be measured on a given day. If that (low) median were compared to the reference median the conclusion might be that the statement "there is no effect" (i.e., the null hypothesis) cannot be rejected. This conclusion is incorrect since we "know" that the effluent produces an elevated chemical concentration. This is called a Type II error (i.e., false negative; usually designated by the Greek letter beta; ß) and leads an environmental manager to the incorrect conclusion that there is no change in chemical concentration due to the effluent.

Another type of error that can arise when testing hypotheses is a Type I error (i.e. false positive; usually designated by the Greek letter alpha; α) arising when a particular dataset leads to the incorrect conclusion that (in this example) there is a significant elevation in chemical concentration ("there is an effect").

There are currently no specific legislated Type I and II error rates. Some authors have suggested that error rates should be no more than 20%; Environment Canada (2002; 2004) discusses error rates of 10%. There is precedence in the north for setting the

Type II error rates at this level. Environment Canada (2002; 2004) advocates setting the Type I error rate equal to the Type II error rate which reflects equal risk to the environment and dischargers. Given the importance of this topic, error rates should be discussed when an AEMP is designed and the Type I error rate must be specified before data are statistically analysed. Otherwise, the validity of any conclusions reached is suspect.

7.3.2 Statistical Power and Environmental Decision Making

The Type I and II errors discussed above are summarized in Table 4 in preparation for a discussion regarding statistical power and environmental decision making. If a Type II error is designated as ß, then the statistical power of a hypothesis test is 1 - ß (for a given effect size). Statistical power is the probability that the statistical test will correctly reject the null hypothesis. It is important to have a high statistical power so that incorrect decisions are not made. The power achieved by a statistical test can be estimated after data are collected. Now consider the case where a statistical null hypothesis is not rejected; one common form of AEMP interpretation is:

There is no environmental effect.

This is not really what the hypothesis test is telling us. A more correct statement is:

There is insufficient evidence to reject the null hypothesis.

One might infer from the first statement that there is no environmental effect (as measured by the endpoint being tested); however, consider the case where only a few samples were collected and the statistical power of the test conducted is found to be low (less than 80%). This means that the ability of the statistical test to correctly reject the null hypothesis is only 80%. There is a 20% chance of stating that there is an effect when in reality there is no effect. With a type I error rate = 10%, there is only a 10% chance that a hypothesis test will incorrectly conclude that there is an effect. Since we want to protect the environment to at least the same degree that a proponent is protected, the most meaningful conclusion when a null hypothesis is not rejected and the statistical power is low is:

The data are inconclusive.

Compare the statement above with the first statement: "There is no environmental effect." This latter statement gives an assurance that no environmental effects are observed, but a more correct statement is that the data are inconclusive. Statistical power allows us to improve our decision making by discriminating between the situation where the data are confirming that there is no environmental effect and an inability to detect an environmental effect because of insufficient statistical power. Statistical power varies with:

- **Type I Error Rate**: For a given sample size, as the Type I error rate gets smaller (90% to 80% - protecting the discharger), then statistical power goes down (decreased environmental protection).
- **Sample Size**: If the Type I error rate is fixed and the sample size increased, then statistical power increases (increased environmental protection).
- **Effect Size**: As the effect size increases and sample size is held fixed, the statistical power increases (i.e., it is easier to detect a large change than it is to detect a small change).
- **Variability**: If the data are highly variable, it is more difficult to detect a change. Statistical power is inversely related to variability. Variability can be reduced by increasing sample sizes, by enhancing quality assurance, and by controlling confounding variables at either the sample collection stage or at the data analysis stage by using statistical tools.

In summary, statistical power is a critical component of any monitoring design where the absence of a demonstrable effect is a desirable outcome. Statistical power may be estimated following initial data collection and if statistical power is unacceptably low (i.e. less than 80%), subsequent sampling may require an increased number of samples and/or there must be better control of variability. When the statistical power is = 80% and the null hypothesis is not rejected, the most appropriate statement is that the hypothesis test is inconclusive. Another way of saying this in the context of an AEMP hypothesis test for a VEC is that:

We could not tell if there was an effect on the VEC.

7.4 Other Monitoring Design Considerations

Whenever opportunities arise that would allow for coupling or adding to the AEMP program and AEMP results, these opportunities should be identified, evaluated, and exploited, to the extent possible. Throughout the course of operations there will be a number of other ways to enhance the effectiveness and efficiency of the AEMP. Three of such opportunities are briefly discussed in this section.

7.4.1 Cooperation

Monitoring designs should take advantage of other data collected by a proponent. In Canada, effluent compliance monitoring is required at the end of pipe for metal mines as part of the EEM program as per the *Metal Mining Effluent Regulations*. When an AEMP is also required all efforts should be made to ensure both program requirements are coordinated. At times, various special studies are conducted by a proponent. Such studies should be viewed as an opportunity by the proponent to 1) reduce monitoring costs if possible by using samples for multiple purposes; 2) increase the scope over which conclusions are made; and, 3) to calibrate various sampling methodologies against one another for additional monitoring cost reductions and/or corroboration of non standard methods.

7.4.2 Synoptic Sampling

An AEMP will require measurement of several VECs. Synoptic measurements (measurements taken at the same time) generally provide a more compelling conclusion than data collected at different times. Also, synoptic measurements allow for more detailed investigations of the data collected at a minimal additional cost. Such sampling and analyses can reduce sample size requirements by controlling for extraneous sources of variability. VECs should be sampled synoptically unless strong reasons for doing otherwise are presented.

7.4.3 Historical Data

Historical data are available only infrequently in the NWT. However, when historical data are available, consideration should be given to previously sampled areas (researchers and government agencies collect information throughout the North). The historical data may also be used for initial sample size estimates.

8.0 Evaluation and Refinement of Aquatic Effects Monitoring Programs

In the early stages of a development, there may be little or even no, site specific information available for a study area. As such, AEMPs are created based on the available TK-based and western science-based information, the available scientific guidance, and the decisions made by interested parties and regulatory boards. However, adaptations to the AEMP may be required after the first year of sampling (if results warrant), but most importantly following three years of monitoring and results. Additionally, an AEMP should be subject to comprehensive review if:

- If Action Levels are exceeded;
- If substantive changes occur in the stressor being monitored. This may include changes in the volume or composition of effluent produced, the manner in which effluent is discharged to the environment, and others;

- If substantive changes occur in the receiving environment (dams, additional discharges in the local receiving environment, etc.);
- In response to changing social values;
- In response to changing regulatory requirements; and/or,
- To take advantage of substantive improvements in monitoring program design, implementation, or interpretation tools.

9.0 Summary

Monitoring designs are a critical part of an AEMP. The primary task when selecting a monitoring design is understanding the physical nature of the immediate receiving environment and any influences from developments. This understanding will enable the selection of exposure and reference sites so as to minimize the effect of confounding variables. The selection of reference sites at the local scale is preferred because they support utilization of a control impact or gradient design. When a suitable reference area is available, the control impact design (including BACI designs) is recommended due to its simplicity and low relative cost. In the event that local reference sites are unavailable, confounding variables at the regional scale must be addressed to create a pool of reference stations that provide the basis for comparison for the exposure sites.

Aside from the conclusions reached following data collection and interpretation, the monitoring design is most often the focus of concerns and doubts expressed regarding conclusions. However important the physical selection of sample locations is, other elements of AEMP design are equally important. These elements start with problem formulation and conceptual site model design and include quality assurance, sample collection protocols, and chemical analysis protocols. Some of the elements, such as effect sizes, VECs and degrees of acceptable uncertainty, are a blend of ecological, social and political values. It is important that Aboriginal governments/organizations, federal and territorial governments, regulatory boards and other interested parties enter the AEMP design process at an early stage and are not only aware of the implications of decisions, such as choice of VEC, but also less familiar concepts, such as statistical power and effect sizes.

As a whole an AEMP is a collection of processes (sample collection, analyte measurement, data interpretation, etc.) and decisions (VECs, acceptable error rates, etc.) meant to inform environmental decision makers. Any monitoring program that does not enable decision makers to make environmental management decision has not met its objectives.

This Technical Guidance Document outlines one part of the process for developing efficient, effective, meaningful, and defensible AEMPs given our current state of knowledge regarding environmental monitoring in the NWT. Future versions of this guidance will continue to evolve to reflect increased understanding of ecological conditions in the North, improved monitoring designs, changes in development technologies, increased development and social valuations.

10.0 References Cited

Angermeier P.L. and R.A. Smogor. 1995. Estimating number of species and relative abundances in stream fish communities: effects of sampling effort and discontinuous spatial distributions. Canadian Journal of Fisheries and Aquatic Sciences 52:936 949.

Bailey, R.C., R.H. Norris, T.B. Reynoldson. 2003. Bioassessment of Freshwater Ecosystems Using the Reference Condition Approach. Kluwer Academic Publishers. Boston.

Blocksom, K.A. and J.E. Flotemersch. 2005. Comparison of macroinvertebrate sampling methods for non wadeable streams. Environmental Monitoring and Assessment 102:243 262.

Busch, W.D.N. and P.G. Sly. (Eds.) 1992. The development of an aquatic habitat classification system for lakes. CRC Press . Boca Raton.

Conquest, L.L., S.C. Ralph and R.J. Naiman. 1994. Implementation of large scale stream monitoring efforts: sampling design and data analysis issues. Pp. 69 90. In Biological Monitoring of Aquatic Systems (Loeb, S.L. and A. Spacie, Eds.). Lewis Publishers. Boca Raton, Florida.

Corkum, L.D. 1992. Spatial distributional patterns of macroinvertebrates along rivers within and among biomes. Hydrobiologia. 239:101 114.

Cott, P.A., A.M. Gordon, K.J. Bourassa, M. Lange and B.A. Zajdlik. In press. Effects of forest fire on young of the year northern pike in the Northwest Territories. Canadian Technical Report of Fisheries and Aquatic Sciences. In press.

Cowardin, L.M., V. Carter, F.C. Golet and E.T. LaRoe. 1979. Classification of wetlands and deepwater habitats of the United States. United States Department of the Interior. Fish and Wildlife Service. Washington, District of Columbia. Jamestown, North Dakota: Northern Prairie Wildlife Research Center Home Page: http://www.npwrc.usgs.gov/resource/1998/classwet/ classwet.htm (Version 04DEC98).

Environment Canada. 2002. Metal mining guidance document for aquatic environmental effects monitoring, June 2002. National Environmental Effects Monitoring Office. Gatineau,Quebec.

Environment Canada. 2003. Revised technical guidance on how to conduct effluent plume delineation. March 2003. National Environmental Effects Monitoring Office. Gatineau,Quebec.

Environment Canada. 2004. Pulp and Paper EEM guidance document. May 2004. National Environmental Effects Monitoring Office. Gatineau,Quebec

Environment Canada. 2008. Understanding the pulp and paper environmental effects monitoring program. http://www.ec.gc.ca/eem/English/Publications/web_publication/EEM101/EEM101.cfm

Flotemersch, J.E. and K.A. Blocksom. 2005. Electrofishing in boatable rivers: Does sampling design affect bioassessment metrics? Environmental Monitoring and Assessment 102:263 283.

Flotemersch, J.E., J.B. Stribling and M.J. Paul. 2006. Concepts and approaches for the bioassessment of non wadeable streams and rivers. EPA/600/R 06/127. United States Environmental Protection Agency. Cincinnati, Ohio.

Frissell, C.A., W.J. Liss, C.E. Warren and M.D. Hurley. 1986. A hierarchical framework for stream habitat classification, viewing streams in a watershed context. Environmental Management 10:199 214.

Gorman, O.T. and J.R. Karr. 1978. Habitat structure and stream fish communities. Ecology 59:507 515.

Green, R.H. 1979. Sampling design and statistical methods for environmental biologists. Wiley Interscience. Chichester, England.

Hawkins, C.P. and M.R. Vinson. 2000. Weak correspondence between landscape classifications and stream invertebrate assemblages: Implications for bioassessment. Journal of the North American Benthological Society 19(3):501 517.

Hawkins, C.P., R.H. Norris, J. Gerritsen, R.M. Hughes, S.K. Jackson, R.K. Johnson and R.J. Stevenson. 2000. Evaluation of the use of landscape classifications for the prediction of freshwater biota: synthesis and recommendations. Journal of the North American Benthological Society 19:541 556.

Heino, J., H. Mykrä and J. Kotanen. 2008. Weak relationships between landscape characteristics and multiple facets of stream macroinvertebrate biodiversity in a boreal drainage basin. Landscape Ecology 23:417 426.

Hughes, R.M., S.A. Heiskary, W.J., Matthews and C.O. Yoder. 1994. Use of ecoregions in biological monitoring. In Biological Monitoring of Aquatic Systems (Loeb, S.L. and A. Spacie, Eds.). pp. 125 149. Lewis Publishers. Boca Raton, Florida.

Hurlbert, S.H. 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monographs 54:187 211.

Imhof, J.G., J. Fitzgibbon, and W.K. Annable. 1996. A hierarchical evaluation system for characterizing watershed ecosystems for fish habitat. Canadian Journal of Fisheries and Aquatic Sciences 53(S1):312 326.

Kokelj S.V., B. Zajdlik and M.S. Thompson. 2009. The impacts of thawing permafrost on the chemistry of lakes across the subarctic boreal tundra transition, Mackenzie Delta region, Canada. Permafrost and Periglacial Processes 20:1-15.

Leopold, L.B. 1994. A View of the River. Harvard University Press. Cambridge, Massachusetts.

Lyons, J. 1992. The length of stream to sample with a towed electrofishing unit when fish species richness is estimated. North American Journal of Fisheries Management 12:198 203.

Mapstone, B.D. 1995. Scalable decision rules for environmental impact studies: effect size, Type I and Type II errors. Journal of Applied Ecology 5:401 410.

Marshall, I. B. and P. H. Schut. 1999. A national ecological framework for Canada. A cooperative product by Ecosystems Science Directorate. Environment Canada and Research Branch. Agriculture and Agri Food Canada. http://sis.agr.gc.ca/cansis/nsdb/ecostrat/intro.html.

Maxwell, J.R., C.J. Edwards, M.E. Jensen, S.J. Paustian, H. Parrott, and D.M. Hill. 1995. A hierarchical framework of aquatic ecological units in North America (Nearctic Zone). General Technical Report NC 176. St. Paul, Minnesota. United States Department of Agriculture. Forest Service. North Central Forest Experimental Station.

MDEQ (Michigan Department of Environmental Quality). 1997. Revision. GLEAS Procedure 51 Qualitative Biological and Habitat Survey Protocols for Wadeable Stream and Rivers. Michigan Department of Environmental Quality.

Newbury, R.W. 1984. Hydrological determinants of aquatic insect habitats. In: The Ecology of Aquatic Insects (Resh, V.H and D.M. Rosenberg, Eds.). pp. 323 357. Praeger, New York.

Ohio EPA (Environmental Protection Agency). 2006. The qualitative habitat evaluation index (QHEI): rationale, methods, and application. Ohio EPA Technical Bulletin EAS/2006 06 1.

Patton, T.M., W.A. Hubert, F.J. Rahel and K. G. Gerow. 2000. Effort needed to estimate species richness in small streams on the Great Plains in Wyoming. North American Journal of Fisheries Management 20:394 398.

Pienitz, R. J.P. Smol, and D.R.S. Lean. 1997a. Physical and chemical limnology of 59 lakes located between the southern Yukon and the Tuktoyaktuk Peninsula, Northwest Territories (Canada). Canadian Journal of Fisheries and Aquatic Sciences 54:330 346.

Pienitz, R. J.P. Smol, and D.R.S. Lean. 1997b. Physical and chemical limnology of 24 lakes located between Yellowknife and Contwoyto Lake, Northwest Territories (Canada). Canadian Journal of Fisheries and Aquatic Sciences 54:347 358.

Reynoldson, T.B., C. Logan, T. Pascoe, and S.P. Thompson. 2003. Canadian Aquatic Biomonitoring Network (CABIN) invertebrate biomonitoring field and laboratory manual. National Water Research Institute. Environment Canada. Saskatoon, Saskatchewan.

Sandin L. and R.K. Johnson. 2000. Ecoregions and benthic macroinvertebrate assemblages of Swedish streams. Journal of the North American Benthological Society 19:462 474.

Smith, E.P. 2002. BACI design. Encyclopedia of Environmetrics (ISBN 0471 899976) Volume 1: 141-148. Edited by A.H. El-Shaarawi and W.W. Piegorsch. West Sussex, United Kingdom.

Sokal, R. and J. Rolf. 1995. Biometry. The principles and practice of statistics in biological research. W. A. Freeman and Company. San Francisco.

Steel, R.G.D., J.H. Torrie and D.A. Dickey. 1997. Principles and Procedures of Statistics. 3rd ed. McGraw Hill, New York.

Stewart Oaten, A. and J.R. Bence. 2001. Temporal and spatial variation in environmental assessment. Ecological Monographs 71:305 339.

Stewart Oaten, A., W.W. Murdoch and K. R. Parker. 1986. Environmental Impact Assessment: APseudoreplication@ in Time? Ecology 67:929 940.

Strahler, A.N. 1957. Quantitative analysis of watershed geomorphology. Transactions of the American Geophysical Union 38:913 920.

Underwood, A.J. 1991. Beyond BACI: Experimental designs for detecting human environmental impacts on temporal variations in natural populations. Australian Journal of Marine and Freshwater Research 42:569 587.

Underwood, A.J. 1994. On beyond BACI: Sampling designs that might reliably detect environmental disturbances. Ecological Applications 4(1):3 15.

USEPA (United States Environmental Protection Agency). 1998. Lake and reservoir bioassessment and biocriteria Technical guidance document. EPA 841 B 98 007. Office of Water. Washington, District of Columbia.

Van Sickle J. and R.M. Hughes. 2000. Classification strengths of ecoregions, catchments, and geographic clusters for aquatic vertebrates in Oregon. Journal of the North American Benthological Society 19:370 384.

Vannote, R.L., G.W. Minshall, K.W. Cummins, J.R. Sedell and C.E. Cushing. 1980. The river continuum concept. Canadian Journal of Fisheries and Aquatic Sciences 37:130 137.

Voelz, N.J. and J.V. McArthur. 2000. An exploration of factors influencing lotic insect species richness. Biodiversity and Conservation 9:1543 1570.

Wetzel, R.G. 1975. Limnology. W.B. Saunders Co. Philadelphia, Pennsylvannia.

Wiens, J.A. 2002. Riverine landscapes: taking landscape ecology into the water. Freshwater Biology 47:501 515.

Wilhelm, J.G.O., J.D. Allan, K.J. Wessell, R.W. Merritt and K. W. Cummins. 2005. Habitat assessment of non wadeable rivers in Michigan. Environmental Management 36:592 609.

Zar, J.H. 1999. Biostatistical Analysis, Fourth Edition. Prentice Hall, New Jersey.

Tables

Table 1. Advantages and disadvantages of monitoring program designs.

Design	Advantage	Disadvantage
Control Impact (includes BACI, BACIP, etc.)	- Easily understood - Easily interpreted - Lower relative cost	- Requires a suitable reference area. - Requires a relatively shallow exposure gradient.
Simple Gradient	- Easily understood - Lower relative cost - Does not require a reference area	- Requires a relatively steep effect gradient to detect effects. - Can be difficult to avoid confounding factors.
Multiple Gradient	- Deals with confounding variables - Does not require a reference area	- Requires a relatively steep effect gradient to detect effects. - Difficult to interpret - Difficult to understand - Higher relative cost
Regional Reference Design	- Does not require a reference area - Intuitive at the conceptual level	- Difficult to interpret - Difficult to understand - Highest relative cost

Table 2. Environment Canada (2008) recommended effect sizes.

Fish Populations[1]		Benthic Macroinvertebrate Communities	
Measurement Endpoint	Effect Size[2]	Measurement Endpoint	Effect Size[3]
Relative Gonad size	25%	Abundance	± 2SD
Relative Liver size	25%	Richness	± 2SD
Condition	10%	Simpson's Eveness	± 2SD
		Abundance	± 2SD

[1] Effect sizes for fish survival are not presented because "Environment Canada is confident that important effects will be detected using the effect sizes in Table 3 because changes in survival of adult fish populations is usually preceded by changes in reproduction" (Environment Canada 2008).

[2] % relative to reference mean.

[3] Standard deviations of reference area.

Table 3. Sample sizes (from Environment Canada 2004).

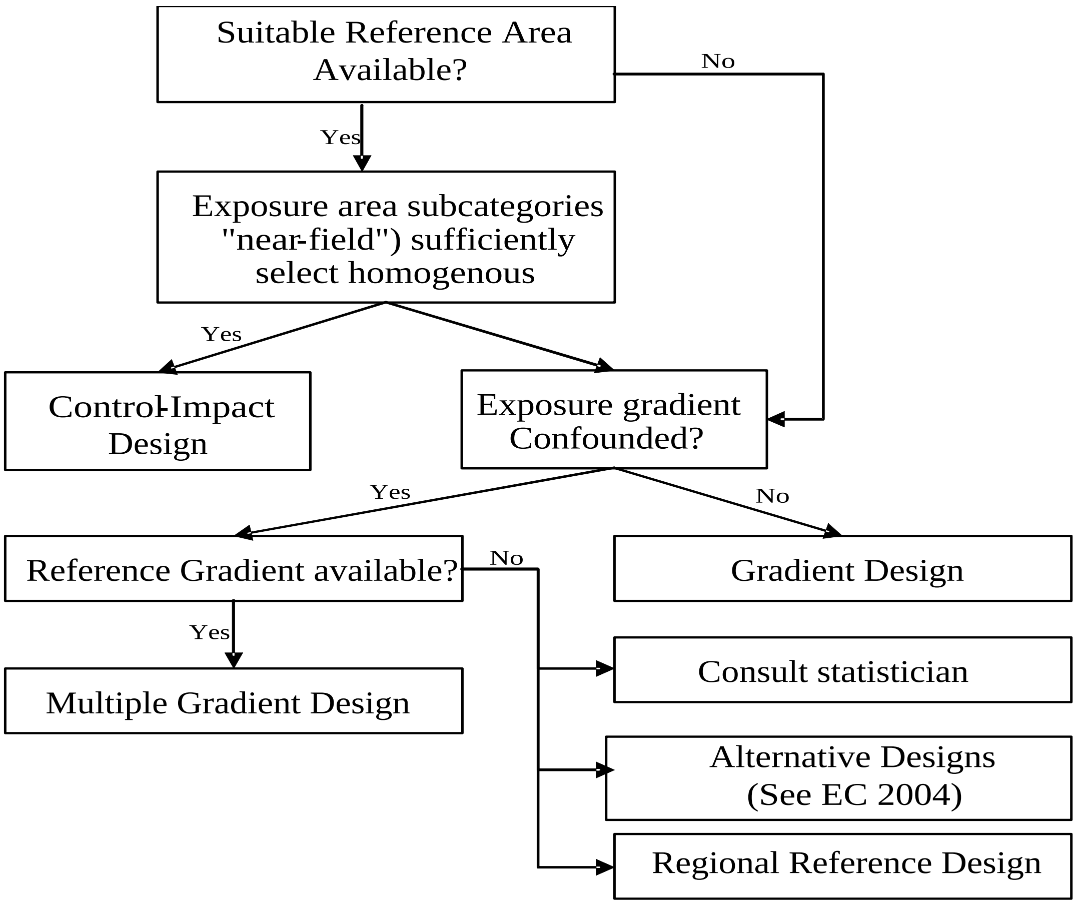
	Statistical Power (%)			
Type I Error Rate (%)	**99 (or 1% Type II error rate)**	**95 (or 5% Type II error rate)**	**90 (or 10% Type II error rate)**	**80 (or 20% Type II error rate)**
1	14	11	10	8
5	11	8	7	5
10	9	7	5	4

Table 4. Relationships between Type I and II errors.

	Null Hypothesis is Correct	**Null Hypothesis is Incorrect**
Do not reject Null Hypothesis	Correct inference	Type II Error (β)
Reject Null Hypothesis	Type I Error (α)	Correct Inference

Figure

Figure 1. Monitoring design selection flowchart.



Appendix

Appendix 1 Small Scale Selection of Sampling Locations

Careful selection of sample locations at a small scale is critical for a defensible AEMP (USEPA 1998). Corkum (1992) investigated variability in benthic macroinvertebrate community composition across longitudinal gradients in three geographic areas of Canada and concluded that site specific factors such as local riparian vegetation and land use were more important than longitudinal gradients in determining benthic macroinvertebrate composition. Other researchers have examined the relative contributions of regional versus local scale variation from the opposite perspective; (i.e., do regional scale classifications capture substantive portions of benthic macroinvertebrate community variability?).

Heino *et al.* (2008) studied the utility of catchment area characteristics and diversity of benthic macroinvertebrates in boreal forests. They concluded that: “landscape catchment variables may have only limited value in explaining variability in biodiversity indices, at least in regions with no strong anthropogenic gradients in land use.” Hawkins and Vinson (2000) found that the landscape classifications catchment size, ecoregion, and life zone explained at best only 14% of the mean difference between within and among classification variability in benthic macroinvertebrate assemblages. Sandin and Johnson (2000) studied variability in benthic macroinvertebrate assemblages at the scale of ecoregions across Sweden and concluded that: “Ecoregion classifications alone, therefore, may not sufficiently partition variance in assemblage composition, and a nested approach, including other factors such as altitude, stream size, and catchment characteristics, is probably needed to improve ecoregion classifications and biological assessments that use stream benthic macroinvertebrates.” With respect to vertebrates, Van Sickle and Hughes (2000) found that regional classifications are useful in classifying assemblages (although they speculate that the observed utility may be due autocorrelation) but do not capture substantive proportions of variability.

Regional variation in surficial water quality also exists within the Northwest Territories. Pienitz *et al.* (1997a) showed that in the north western portion of the Northwest Territories, there were systematic variations in ionic composition with Ca-Cl-Na predominating in waters near the Arctic Ocean, with Ca- HCO_3 increasing with distance from the sea. Further inland, Pienitz *et al.* (1997b) found that nutrients and major ions increased with decreasing latitude to the conifer catchments in the south. Kokelj *et al.* (2009) investigated 73 small lakes in the Mackenzie Delta area and found substantive differences in water quality as a function permafrost degradation with secondary influences of surrounding burns and lake area.

USEPA (1998) acknowledges the effect of ecoregions on both aquatic biota and water quality. They recommend a hierarchical approach to site selection beginning with ecoregion, possibly followed by watershed characteristics (lake drainage type, surrounding land use, the watershed to lake area ratio, slope and surrounding geology), specific lake basin characteristics (surface area to volume ratio, bottom type, mean and maximum depth), specific lake hydrological characteristics (retention time, stratification, circulation and water level fluctuations) and specific lake water chemistry variables (alkalinity, conductivity, turbidity, DIC, DOC and colour). The following sections discuss small-scale selection criteria in general, and specifically in the context of lotic and lentic receiving environments. Criteria specific to a valued ecosystem component (VEC) are presented within sub sections. Overall, this information is intended to support the identification of appropriate sampling sites for inclusion in a monitoring program.

A1.1 General

A1.1.1 Habitat Classification

Environment Canada (2004) recommends the habitat classification scheme described in Cowardin *et al.* (1979) and Busch and Sly (1992), but only to the "class" level. The criteria within this classification scheme are nested in a hierarchical arrangement of aquatic systems (their Figure 2.1), beginning with the "system" level (marine, river, estuarine and lacustrine), followed by the "subsystem" level, and ending at the recommended "class" level which is substrate type.

Environment Canada (2004) recommends that this classification scheme be augmented by an assessment of thermal and chemical stratification. These should be conducted at times when minimal conditions are expected, for example, low oxygen conditions in late winter. Also, salinity is a criterion, with seven levels as described in Table 2.3 (Environment Canada 2004) for marine, estuarine and brackish environments and water regime-modifying criteria are presented for tidal (4 classes) and non tidal (6 classes) waters. Environment Canada (2004) suggests that the following general criteria should be applied to sample site location:

- Bedrock and surficial geology;
- Topography;
- Soil and vegetation;
- Site accessibility; and,
- Climatology at a relevant scale.

All of these general criteria fall within the purview of habitat classification. If one biotic VEC is of much greater importance then another, habitat classification schemes specific to that VEC might supersede generic habitat classification schemes. For example, Imhof *et al.* (1996) discuss habitat assessment from the perspective of fish alone.

A1.1.2 Exposure

The information on mixing of the effluent in the exposure area is necessary to establish the mixing zone and to define the near field and far field exposure areas. Models may be used to predict effluent dilution at the Conceptual Study Design stage of AEMP development, but must be confirmed by plume delineation studies upon project commencement. The plume delineation studies should consider the worst case scenario, which involves a combination of low dilution and possibly thermal stress, timing with respect to critical life stages (fish swim up), and other factors. Plume delineation studies should be repeated when effluent volume changes by more than 25% of the original approved discharged capacity and the point of discharge changes in any substantive manner. Guidance on conducting plume delineation studies is found in Environment Canada (2003).

Within the effluent plume, depositional and scouring zones should be identified as contaminants are often sequestered within depositional zones. In keeping with the criterion for selecting samples with maximum exposure (discussed in Sections A1.1.5 and A1.1.6), these areas might be preferentially sampled. Issues regarding sampling within the exposure area specific to a given monitoring design are discussed in the relevant sub sections.

A1.1.3 Hydrology

The following hydrologic criteria should be applied to site selection: description of the exposure area watershed, bathymetry, tides if applicable, thermal and chemical stratification if applicable, and natural barriers to fish movement (Environment Canada 2004).

A1.1.4 Anthropogenic Influences

The following anthropogenic influences (modified from Environment Canada 2004) should be considered when selecting sample locations:

- Water diversions, such as dams and culverts;
- Boat related facilities such as docks, launches, ferry launch ramps, and fuelling stations;
- Land transport structures such as bridges and fords;
- Discrete inputs to the waterbody such as effluent, stormwater or sewage outfalls; and,
- Diffuse inputs to the waterbody due to contaminated groundwater movement, aerial plume fallout, and other inputs.

A1.1.5 Benthic Macroinvertebrates

Environment Canada (2004) makes the following recommendations for choosing sampling areas for benthic macroinvertebrates. Choose sample locations:

- In the predominant habitat, in the area most exposed to effluent, exhibiting the greatest diversity; and,
- Giving consideration to previously sampled areas and confounding factors.

They also recommend that selection of the sample season should consider:

- Sampling season used in previous studies;
- Season when diversity is highest; and,
- Season when exposure to effluent (stressor) is highest.

A1.1.6 Fish

The following criteria (modified from Environment Canada 2004) for choosing sampling areas for fish should be considered:

- Traditional Knowledge (TK):
  - Consultation with Aboriginal governments/organizations provides a means of acquiring TK on fish distribution and abundance.

- General Factors Influencing Fish or Fish Habitat:
  - Consideration of structures that can influence fish movement, such as dams, waterfalls and culverts; and,
  - Commercial, recreational or Aboriginal fisheries.

- Ecological;
  - Use of habitat for purposes such as spawning, early rearing, foraging, and other activities;
  - Presence of rare species; and,
  - Proportions of substrate types; depositional and erosional zones, and macrophyte cover.

- Stressor Related;
  - Areas where fish are exposed to the highest effluent (stressor) concentration and areas further removed to establish geographic extent of effects, if any;

- Abundance of selected species in the area being sampled to gauge possible sampling effects;
- Areas previously sampled; and,
- Confounding factors.

A1.2 Lotic Receiving Environments

In addition to the general criteria described in Section A.1.1, the following criteria should be addressed when monitoring lotic environments (i.e., flowing waters). These criteria are broken down into the somewhat overlapping sections: general; habitat classification; and, geomorphology.

A1.2.1 General Considerations

- Factors relevant to sampling reach length decisions in wadeable streams (e.g., Patton *et al.* 2000; Lyons 1992)
- Flotemersch *et al.* (2006; Table 4-1) discuss three large river characterization protocols for evaluation of habitat to understand biota.

A1.2.2 Habitat Classification Considerations

A large proportion of the variability in biological diversity in rivers is explained by diversity in habitat (Gorman and Karr 1978; Vannote *et al.* 1980; Voelz and McArthur 2000; Flotemersch *et al.* 2006). Therefore, it is critical to correctly assess habitat in order to prevent habitat differences from obscuring stressor related effects or inducing apparent stressor related effects. The level of sampling effort needed to characterize habitat for the purpose of assessing potential confounding variables for effects on biological variables is less than conducting a survey to define ecological condition. Environment Canada (2004) suggests that habitat assessment should address the following (in addition to those criteria presented in Section A.1.1):

- Elevation and gradient;
- Any dams, falls or other barriers to fish migration;
- Ranges and mean annual discharge;
- General substrate characteristics (preferably in the form of a gradient profile chart); and,
- Presence of any upstream and downstream inputs (e.g., storm water, sewer overflow, effluent from other industrial sites).

Flotemersch *et al.* (2006) discuss the bioassessment of large non wadeable rivers. In that document, a variety of protocols are reviewed and criteria for determining sample locations are presented. Of the protocols presented, the less intensive habitat assessment protocols designed to support biological assessments, rather than those for conducting complete ecological assessments, should be reviewed. These include:

- The Large River Bioassessment Protocol (Blocksom and Flotemersch 2005; Flotemersch and Blocksom 2005);
- The Non Wadeable Stream Habitat Index (MDEQ 1997 and Wilhelm 2005); and,
- The Qualitative Habitat Evaluation Index (Ohio EPA 2006).

A1.2.3 Geomorphology Considerations

A1.2.3.1 Stream Order

This is the position of stream within the hierarchy of the drainage basin (Strahler 1957). Consideration of stream order is recommended by Environment Canada (2004) and Flotemersch *et al.* (2006) when monitoring lotic receiving environments.

A1.2.3.2 Valley Segment

Valley segments are described by Environment Canada (2004) as portions of a drainage basin with the distinctive characteristics, as defined by a combination of flow and geomorphic properties. These characteristics include (from Conquest *et al.* 1994, as cited in Environment Canada 2004):

- Stream order;
- Valley slideslope gradient;
- Ratio of valley bottom width to active channel width;
- Channel gradient;
- Stream corridor geomorphic surface deposits; and,
- Channel pattern.

Consideration of valley segment is recommended by Environment Canada (2004) when monitoring lotic receiving environments.

A1.2.3.3 Reaches

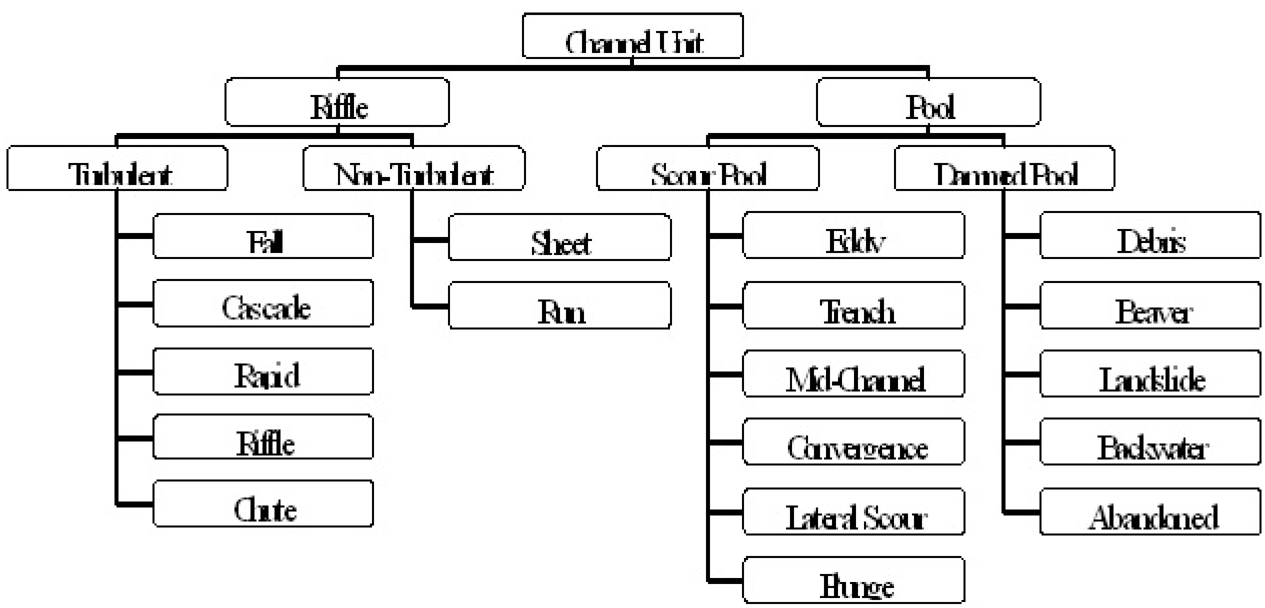
A stream reach is a homogenous portion of a river or stream as defined by various sets of criteria (Newbury 1984; Frissell *et al.* 1986 and Leopold 1994). A set of criteria most relevant to the NWT should be defined by aquatic biologists, geologists, hydrogeologists and interested parties. One criterion that may be important in a low lying area of a large river, such as the delta area of the Mackenzie River, is the meander length. Meander lengths are discussed in Flotemersch *et al.* (2006).

A1.2.3.4 Channel Units

Channel units are the level at which individual samples are collected. Channel units are defined by micro scale criteria within the hierarchy of ecoregion -> watershed -> -> valley segment -> reaches. (Environment Canada 2004)

A1.3 Lentic Receiving Environments

In addition to the general criteria described in Section A.1.1, the following criteria should be addressed when monitoring lentic environments (i.e., lakes, ponds, etc). These criteria are broken down into the somewhat overlapping sections, including habitat classification, and geomorphology.



A1.3.1 Habitat

Habitat assessment should address the following (Environment Canada 2004):

- Bathymetry;
- The locations of major inlets and outlets;
- General oxygen-temperature conditions (e.g. thermal stratification, occurrences of oxygen depletion in deep water); and,
- Dispersion (Environment Canada 2004; Table 2.1).

The criteria discussed for selecting lentic reference sites (Section A1.3) are also suitable for identification of exposure sites.

A1.3.2 Geology

The geology of a lake affects the biology and chemistry of a lake (Frissel *et al.* 1986; Hawkins *et al.* 2000; Wiens 2002). Environment Canada (2004) recommends that geology be considered when selecting reference lakes.

A1.3.3 Morphometry

Basin morphometry can influence productivity and biological assemblages due to size of epilimnetic zone, residence time, and thermal and/or chemical stratification. Environment Canada (2004) recommends that lake morphometry be considered when selecting reference lakes.

A1.3.4 Trophic Status

The trophic status of many arctic and alpine tundra lakes in the North is oligotrophic to ultraoligotrophic, whereas lakes in the boreal forest and along the forest-tundra boundary are oligotrophic to mesotrophic (Pienitz *et al.* 1997a). One very real concern among Aboriginal governments and organizations in the NWT is a change in water quality due to nutrient inputs from industrial developments. Given this, special care should be taken to accurately and precisely define trophic status of arctic and alpine tundra lakes in the North. Environment Canada (2004) recommends that trophic status be considered when selecting reference lakes.

A1.3.5 Lake Zone

Biological assemblages and chemical characteristics vary by zone within a lake (Wetzel 1975) and, therefore, lake zone must be considered when selecting sampling locations. Lakes include both a littoral zone (region from upper spray limit to depth of rooted plants), and a profundal zone (section beyond depth of rooted plants). Environment Canada (2004) recommends that lake zone be considered when selecting reference lakes.

A1.4 Coastlines

The definition of "open coastline" is unclear; however Environment Canada classifies the Great Lakes as open coastlines. In the North, the shorelines of Great Slave and Great Bear Lakes might be classified as "open coastline". Additional criteria following Environment Canada (2004) for evaluating sampling locations along open coastlines are:

- Depth contours;
- Near-shore substrate characteristics;
- Shoreline configuration;
- Locations of inflowing rivers; and,
- Location of discharges and activities that may affect the sampling area.

A1.5 Estuaries

Environment Canada (2004) recommends the following criteria in addition to those described as general criteria when choosing sample locations in an estuary:

- Salinity gradients;
- Flow patterns; and
- Bathymetry.

A1.6 References Cited

Angermeier P.L. and R.A. Smogor. 1995. Estimating number of species and relative abundances in stream fish communities: effects of sampling effort and discontinuous spatial distributions. Canadian Journal of Fisheries and Aquatic Sciences 52:936 949.

Bailey, R.C., R.H. Norris, T.B. Reynoldson. 2003. Bioassessment of Freshwater Ecosystems Using the Reference Condition Approach. Kluwer Academic Publishers. Boston.

Blocksom, K.A. and J.E. Flotemersch. 2005. Comparison of macroinvertebrate sampling methods for non wadeable streams. Environmental Monitoring and Assessment 102:243 262.

Busch, W.D.N. and P.G. Sly. (Eds.) 1992. The development of an aquatic habitat classification system for lakes. CRC Press . Boca Raton.

Conquest, L.L., S.C. Ralph and R.J. Naiman. 1994. Implementation of large scale stream monitoring efforts: sampling design and data analysis issues. Pp. 69 90. In Biological Monitoring of Aquatic Systems (Loeb, S.L. and A. Spacie, Eds.). Lewis Publishers. Boca Raton, Florida.

Corkum, L.D. 1992. Spatial distributional patterns of macroinvertebrates along rivers within and among biomes. Hydrobiologia. 239:101 114.

Cott, P.A., A.M. Gordon, K.J. Bourassa, M. Lange and B.A. Zajdlik. In press. Effects of forest fire on young of the year northern pike in the Northwest Territories. Canadian Technical Report of Fisheries and Aquatic Sciences. In press.

Cowardin, L.M., V. Carter, F.C. Golet and E.T. LaRoe. 1979. Classification of wetlands and deepwater habitats of the United States. United States Department of the Interior. Fish and Wildlife Service. Washington, District of Columbia. Jamestown, North Dakota: Northern Prairie Wildlife Research Center Home Page: http://www.npwrc.usgs.gov/resource/1998/classwet/ classwet.htm (Version 04DEC98).

Environment Canada. 2002. Metal mining guidance document for aquatic environmental effects monitoring, June 2002. National Environmental Effects Monitoring Office. Gatineau,Quebec.

Environment Canada. 2003. Revised technical guidance on how to conduct effluent plume delineation. March 2003. National Environmental Effects Monitoring Office. Gatineau,Quebec.

Environment Canada. 2004. Pulp and Paper EEM guidance document. May 2004. National Environmental Effects Monitoring Office. Gatineau,Quebec

Environment Canada. 2008. Understanding the pulp and paper environmental effects monitoring program. http://www.ec.gc.ca/eem/English/Publications/ web_publication/EEM101/EEM101.cfm

Flotemersch, J.E. and K.A. Blocksom. 2005. Electrofishing in boatable rivers: Does sampling design affect bioassessment metrics? Environmental Monitoring and Assessment 102:263 283.

Flotemersch, J.E., J.B. Stribling and M.J. Paul. 2006. Concepts and approaches for the bioassessment of non wadeable streams and rivers. EPA/600/R 06/127. United States Environmental Protection Agency. Cincinnati, Ohio.

Frissell, C.A., W.J. Liss, C.E. Warren and M.D. Hurley. 1986. A hierarchical framework for stream habitat classification, viewing streams in a watershed context. Environmental Management 10:199 214.

Gorman, O.T. and J.R. Karr. 1978. Habitat structure and stream fish communities. Ecology 59:507 515.

Green, R.H. 1979. Sampling design and statistical methods for environmental biologists. Wiley Interscience. Chichester, England.

Hawkins, C.P. and M.R. Vinson. 2000. Weak correspondence between landscape classifications and stream invertebrate assemblages: Implications for bioassessment. Journal of the North American Benthological Society 19(3):501 517.

Hawkins, C.P., R.H. Norris, J. Gerritsen, R.M. Hughes, S.K. Jackson, R.K. Johnson and R.J. Stevenson. 2000. Evaluation of the use of landscape classifications for the prediction of freshwater biota: synthesis and recommendations. Journal of the North American Benthological Society 19:541 556.

Heino, J., H. Mykrä and J. Kotanen. 2008. Weak relationships between landscape characteristics and multiple facets of stream macroinvertebrate biodiversity in a boreal drainage basin. Landscape Ecology 23:417 426.

Hughes, R.M., S.A. Heiskary, W.J., Matthews and C.O. Yoder. 1994. Use of ecoregions in biological monitoring. In Biological Monitoring of Aquatic Systems (Loeb, S.L. and A. Spacie, Eds.). pp. 125 149. Lewis Publishers. Boca Raton, Florida.

Hurlbert, S.H. 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monographs 54:187 211.

Imhof, J.G., J. Fitzgibbon, and W.K. Annable. 1996. A hierarchical evaluation system for characterizing watershed ecosystems for fish habitat. Canadian Journal of Fisheries and Aquatic Sciences 53(S1):312 326.

Kokelj S.V., B. Zajdlik and M.S. Thompson. 2009. The impacts of thawing permafrost on the chemistry of lakes across the subarctic boreal tundra transition, Mackenzie Delta region, Canada. Permafrost and Periglacial Processes 20:1-15.

Leopold, L.B. 1994. A View of the River. Harvard University Press. Cambridge, Massachusetts.

Lyons, J. 1992. The length of stream to sample with a towed electrofishing unit when fish species richness is estimated. North American Journal of Fisheries Management 12:198 203.

Mapstone, B.D. 1995. Scalable decision rules for environmental impact studies: effect size, Type I and Type II errors. Journal of Applied Ecology 5:401 410.

Marshall, I. B. and P. H. Schut. 1999. A national ecological framework for Canada. A cooperative product by Ecosystems Science Directorate. Environment Canada and Research Branch. Agriculture and Agri Food Canada. http://sis.agr.gc.ca/cansis/nsdb/ecostrat/intro.html.

Maxwell, J.R., C.J. Edwards, M.E. Jensen, S.J. Paustian, H. Parrott, and D.M. Hill. 1995. A hierarchical framework of aquatic ecological units in North America (Nearctic Zone). General Technical Report NC 176. St. Paul, Minnesota. United States Department of Agriculture. Forest Service. North Central Forest Experimental Station.

MDEQ (Michigan Department of Environmental Quality). 1997. Revision. GLEAS Procedure 51 Qualitative Biological and Habitat Survey Protocols for Wadeable Stream and Rivers. Michigan Department of Environmental Quality.

Newbury, R.W. 1984. Hydrological determinants of aquatic insect habitats. In: The Ecology of Aquatic Insects (Resh, V.H and D.M. Rosenberg, Eds.). pp. 323 357. Praeger, New York.

Ohio EPA (Environmental Protection Agency). 2006. The qualitative habitat evaluation index (QHEI): rationale, methods, and application. Ohio EPA Technical Bulletin EAS/2006 06 1.

Patton, T.M., W.A. Hubert, F.J. Rahel and K. G. Gerow. 2000. Effort needed to estimate species richness in small streams on the Great Plains in Wyoming. North American Journal of Fisheries Management 20:394 398.

Pienitz, R. J.P. Smol, and D.R.S. Lean. 1997a. Physical and chemical limnology of 59 lakes located between the southern Yukon and the Tuktoyaktuk Peninsula, Northwest Territories (Canada). Canadian Journal of Fisheries and Aquatic Sciences 54:330 346.

Pienitz, R. J.P. Smol, and D.R.S. Lean. 1997b. Physical and chemical limnology of 24 lakes located between Yellowknife and Contwoyto Lake, Northwest Territories (Canada). Canadian Journal of Fisheries and Aquatic Sciences 54:347 358.

Reynoldson, T.B., C. Logan, T. Pascoe, and S.P. Thompson. 2003. Canadian Aquatic Biomonitoring Network (CABIN) invertebrate biomonitoring field and laboratory manual. National Water Research Institute. Environment Canada. Saskatoon, Saskatchewan.

Sandin L. and R.K. Johnson. 2000. Ecoregions and benthic macroinvertebrate assemblages of Swedish streams. Journal of the North American Benthological Society 19:462 474.

Smith, E.P. 2002. BACI design. Encyclopedia of Environmetrics (ISBN 0471 899976) Volume 1: 141-148. Edited by A.H. El-Shaarawi and W.W. Piegorsch. West Sussex, United Kingdom.

Sokal, R. and J. Rolf. 1995. Biometry. The principles and practice of statistics in biological research. W. A. Freeman and Company. San Francisco.

Steel, R.G.D., J.H. Torrie and D.A. Dickey. 1997. Principles and Procedures of Statistics. 3rd ed. McGraw Hill, New York.

Stewart Oaten, A. and J.R. Bence. 2001. Temporal and spatial variation in environmental assessment. Ecological Monographs 71:305 339.

Stewart Oaten, A., W.W. Murdoch and K. R. Parker. 1986. Environmental Impact Assessment: APseudoreplication@ in Time? Ecology 67:929 940.

Strahler, A.N. 1957. Quantitative analysis of watershed geomorphology. Transactions of the American Geophysical Union 38:913 920.

Underwood, A.J. 1991. Beyond BACI: Experimental designs for detecting human environmental impacts on temporal variations in natural populations. Australian Journal of Marine and Freshwater Research 42:569 587.

Underwood, A.J. 1994. On beyond BACI: Sampling designs that might reliably detect environmental disturbances. Ecological Applications 4(1):3 15.

USEPA (United States Environmental Protection Agency). 1998. Lake and reservoir bioassessment and biocriteria Technical guidance document. EPA 841 B 98 007. Office of Water. Washington, District of Columbia.

Van Sickle J. and R.M. Hughes. 2000. Classification strengths of ecoregions, catchments, and geographic clusters for aquatic vertebrates in Oregon. Journal of the North American Benthological Society 19:370 384.

Vannote, R.L., G.W. Minshall, K.W. Cummins, J.R. Sedell and C.E. Cushing. 1980. The river continuum concept. Canadian Journal of Fisheries and Aquatic Sciences 37:130 137.

Voelz, N.J. and J.V. McArthur. 2000. An exploration of factors influencing lotic insect species richness. Biodiversity and Conservation 9:1543 1570.

Wetzel, R.G. 1975. Limnology. W.B. Saunders Co. Philadelphia, Pennsylvannia.

Wiens, J.A. 2002. Riverine landscapes: taking landscape ecology into the water. Freshwater Biology 47:501 515.

Wilhelm, J.G.O., J.D. Allan, K.J. Wessell, R.W. Merritt and K. W. Cummins. 2005. Habitat assessment of non wadeable rivers in Michigan. Environmental Management 36:592 609.

Zar, J.H. 1999. Biostatistical Analysis, Fourth Edition. Prentice Hall, New Jersey.