Environmental
Studies
Research
Funds

118 Considerations in the Design of Effects Monitoring Strategies: Beaufort Sea Case Study

Canada

January 1992

ENVIRONMENTAL STUDIES RESEARCH FUNDS

Report No. 118

January 1992

CONSIDERATIONS IN THE DESIGN OF EFFECTS MONITORING STRATEGIES

Beaufort Sea Case Study

D.J. Thomas

Axys Environmental Consulting Ltd.
2045 Mills Road
Sidney, B.C.
V8L 3S8

Scientific Advisor: Dr. D.A. Stone

The correct citation for this report is:

Thomas, D.J. 1992. Considerations in the Design of Effects Monitoring Strategies for the Beaufort Sea. Beaufort Sea Case Study. Environmental Studies Research Funds Report No. 118. Calgary. 54pp.

Published under the auspices
of the Environmental Studies
Research Funds
ISBN 0-921652-19-4

TABLE OF CONTENTS

TABLE OF CONTENTS, cont'd

LIST OF FIGURES

LIST OF TABLES

SUMMARY

In anticipation of future hydrocarbon development in the offshore Beaufort Sea region, there is a clear requirement to conduct monitoring studies that will address the uncertainties associated with environmental impact predictions, assess the effectiveness of mitigative measures, and reassure the public that environmental degradation is not taking place. Most monitoring programs in the past have not met expectations because they have been too ambitious and often have been based on inadequate background information. The results of these studies, therefore, typically produced more questions than answers. The objective of the current study was to develop and describe a systematic framework within which effects-monitoring programs can be designed such that the results obtained will be scientifically sound, relevant and of greater long-term value than has generally been the case in the past.

The interdisciplinary approach outlined emphasises two features -- hypothesis testing and natural variability. Hypothesis testing is the cornerstone of any effects-monitoring program. It focuses the monitoring program by establishing unequivocal monitoring goals, and by identifying the spatial, temporal and resolution scales necessary to test the hypothesis. The importance of spatial and temporal natural variability as a factor in the design of effects-monitoring studies cannot be overemphasised. More than any other single factor, our inability to explain natural variability places a limit on our ability to resolve anthropogenic changes. Without some level of understanding of this matter, most null hypotheses should be left untested. Instead, basic studies should be undertaken to improve our understanding of the natural variability of physical, chemical and biological parameters of the Beaufort Sea Region.

The proposed framework for designing effects-monitoring studies comprises six major stages: (1) identification of activity/receiving environment interactions; (2) quantification of the scale/extent of the disturbance; (3) assessment of possible effects; (4) formulation of the null impact hypothesis; (5) design of the monitoring program; and (6) execution of the monitoring program. The various stages are linked together by feedback loops which permit iteration of the procedures in individual stages of the plan.

The principles of quality assurance / quality control must be applied at each stage. They are the "glue" that will provide the consistency and defensibility of the monitoring program. Although the framework has been developed for the marine environment, it is sufficiently generic to apply to monitoring programs directed at the terrestrial environment such as the Mackenzie Delta.

The key recommendations of this study were as follows:

1. A committee should be identified whose responsibility would be to review and approve monitoring program designs and methods.

2. At the present time, data gaps related to the natural variability of whale, bird and seal populations of the Beaufort Sea prevent proper design of monitoring programs involving these animals. No further effects-monitoring studies should be contemplated until the data gaps are filled.

3. The previous report entitled "Sampling and Analysis in the Arctic Marine Benthic Environment" should be updated to reflect recent advances in protocols particularly for benthic invertebrates.

4. Biological methods should be intercalibrated and steps taken immediately to improve the intercomparability of biological data sets. As biological measurements are more diverse and inherently more variable than chemical measurements, it is unlikely that biological data will ever attain the accuracy of chemical data. Nonetheless, improvements must be made to the way biological measurements are made in order to identify and understand relationships between environmental factors and biological responses.

5. Representatives from government, industry and local native groups should reach an *a priori* agreement on maximum acceptable effects level of environmental disturbance. This will inject a sense of purpose into the process and lead to more effective and meaningful testing of impact hypotheses.

GRAND SOMMAIRE

En prévision du développement futur des forages offshore dans la mer de Beaufort, il est tout à fait indiqué d'effectuer des études de surveillance portant sur des points d'incertitude associés aux prévisions des incidences environnementales, d'évaluer l'efficacité des mesures correctives et de rassurer le public en prenant toutes les mesures nécessaires pour éviter la dégradation de l'environnement. La plupart des programmes de surveillance antérieurs n'ont pas atteint leurs objectifs parce qu'ils étaient trop ambitieux et qu'ils étaient souvent basés sur des informations générales inexactes. Par conséquent, les résultats de ces études produisaient plus de questions que de réponses. L'objectif de la présente étude est d'élaborer et de décrire un cadre systématique à l'intérieur duquel des programmes de surveillance des répercussions peuvent être conçus de façon à ce que les résultats obtenus soient bien fondés scientifiquement, pertinents et qu'ils aient une plus grande valeur à long terme que ceux qu'on a obtenus par le passé, sauf exception.

L'approche interdisciplinaire soulignée comporte deux caractéristiques principales : la vérification des hypothèses et la variabilité naturelle. La vérification des hypothèses est la pierre angulaire de tout programme de surveillance des répercussions. Elle met l'accent sur le programme de surveillance en établissant des objectifs de surveillance non équivoques, et en identifiant les échelles spatiales, temporelles et de résolution nécessaires pour vérifier l'hypothèse. On ne saurait trop insister sur l'importance de la variabilité naturelle spatiale et temporelle comme facteur pour la conception des études de surveillance des répercussions. Plus que tout autre facteur, notre inaptitude à expliquer la variabilité naturelle limite notre aptitude à mesurer les changements anthropiques. Sans une certaine compréhension de cette question, la vérification de la plupart des hypothèses sans incidences est inutile. Il faudrait plutôt entreprendre des études de base afin d'améliorer notre compréhension de la variabilité naturelle et des paramètres physiques, chimiques et biologiques de la région de la mer de Beaufort.

Le cadre proposé pour la conception des études de surveillance des répercussions comprend six étapes principales : 1) identification et interaction entre les activités et l'environnement récepteur; 2) quantification de l'échelle de grandeur et de l'importance de la perturbation; 3) évaluation des effets possibles; 4) formulation de l'hypothèse sans incidences; 5) conception d'un programme de surveillance; et 6) exécution du programme de surveillance. Les différentes étapes sont liées ensemble par des boucles des rétroactions qui permettent l'itération des procédures à des étapes particulières du plan. Les principes de l'assurance de la qualité/contrôle de la qualité doivent être appliqués à chaque étape. Ils constituent le ciment qui assure la cohésion et la robustesse du programme de surveillance. Bien que ce cadre ait été élaboré pour l'environnement marin, il est suffisamment général pour pouvoir être appliqué à des programmes de surveillance destinés à l'environnement terrestre, comme c'est le cas dans le delta du Mackenzie.

Les principales recommandations de l'étude sont les suivantes :

1. Il faut désigner un comité chargé d'examiner et d'approuver les plans et les méthodes de programmes de surveillance.

2. Présentement, à cause de lacunes de données en rapport avec la variabilité naturelle des populations de baleines, d'oiseaux et de phoques de la mer de Beaufort, il est difficile de concevoir des programmes de surveillance adéquats pour ces animaux. On ne devrait pas entreprendre d'autres études de surveillance des répercussions avant que ces lacunes de données n'aient été comblées.

3. Le rapport antérieur intitulé "Sampling and Analysis in the Artic Marine Benthic Environment" (Échantillonnage et analyse de l'environnement benthique marin arctique) doit être remis à jour de façon à refléter les progrès récents dans les protocoles, et plus particulièrement pour les invertébrés benthiques.

4. Les méthodes biologiques doivent être intercalibrées et des mesures doivent être prises immédiatement pour améliorer l'intercomparabilité des ensembles de données biologiques. Étant donné que les mesures biologiques sont plus diverses et, de façon inhérente, plus variables que les mesures chimiques, il est peu probable que les données biologiques atteignent un jour la précision des données chimiques. Néanmoins, il est nécessaire d'améliorer la façon dont les mesures biologiques sont obtenues afin d'identifier et de comprendre les relations entre les facteurs environnementaux et les réponses biologiques.

5. Des représentants du gouvernement, de l'industrie et des groupes autochtones locaux devraient conclure un accord *a priori* sur la valeur maximum acceptable de niveau de répercussions dues à des perturbations environnementales. Ceci introduira un élément directeur dans le processus et permettra d'obtenir une vérification plus efficace et plus significative des hypothèses de répercussions.

INTRODUCTION

INTRODUCTORY NOTE

This report contains general information about designing effects monitoring studies for evaluating environmental effects associated with disturbances to the environment. Although the examples used to illustrate the various elements of the effects monitoring approach are geographically specific to hydrocarbon development in the Beaufort Sea Region, we believe that the concepts, information and tools provided here are sufficiently generic that they can be applied successfully to the development of any monitoring study.

Future hydrocarbon development activities in the offshore Beaufort Sea Region provide a clear need for effective effects monitoring studies. Such studies must (1) address the uncertainties associated with environmental impact predictions; (2) assess the effectiveness of mitigative measures; and (3) re-assure the public that environmental degradation is not taking place. The early attempts at conducting effects monitoring studies in the offshore Beaufort Sea greatly underestimated the difficulty of detecting an impact in a highly variable and poorly understood environment. The failure to anticipate natural variation resulted in the use of study designs that were merely extensions of baseline surveying, the uninformed use of statistical analyses and virtually no dependence upon hypothesis testing. The importance of spatial and temporal natural variability in effects monitoring studies cannot be over-emphasised. Biological systems are inherently noisy. Spatial and temporal heterogeneities in marine ecosystems are a combination of the natural variability of the physical environment and of the biological processes themselves. Natural variability often overshadows impact effects or confounds the resolution of such effects. It must be understood at least on scales above that of the sample size, in order to determine if differences observed in the level of the monitoring variable are attributable to offshore hydrocarbon development activities in the Beaufort Sea. Understanding temporal variability is critical in "before-and-after" comparisons of environmental variables or biological response to impact producing activity. More than any other factor, our inability to explain natural variability places a limit on our ability to resolve anthropogenic changes.

The failure of the early effects monitoring studies to meet expectations was eventually confronted. The early solution, however, was only a marginal improvement. Rather than implement better fundamental study design, modifications were made to many study plans that simply added more stations and more parameters to be measured in the hope that somehow the greater quantity of data would lead to successful studies. This approach, as might be expected, did not work.

A significant improvement to the use of effects monitoring studies in the Beaufort Sea occurred with the initiation of the Beaufort Environmental Monitoring Program (BEMP) in 1983 (Indian and Northern Affairs Canada, 1984 - 1988). The cornerstone of BEMP was the emphasis on impact hypothesis testing using a careful systematic evaluation of linkages that join industry activities and "valued ecosystem components". In general, the result has been a more thoughtful selection of sampling programs, a more knowledgeable use of data analyses and a reduced reliance on unsubstantiated explanation to reconcile study results such that some recent effects monitoring studies have been more focused and practical than were their predecessors.

A continuing shortcoming of many effects studies in the Beaufort Sea, however, is the lack of commitment to careful design and planning. Consequently the objective of this project was to develop a strategy that will ensure, to the extent possible, that future effects monitoring programs related to hydrocarbon development in the Beaufort Sea are designed, conducted, analysed and reported in a manner which is scientifically defensible, relevant, and which will allow confidence to be placed on their results and conclusions. Results must be believable if they are to be used effectively.

As part of the study, a small group of experts in the fields of chemistry, biology and (geo)physics of the Beaufort Sea region was convened in a workshop setting to discuss matters relevant to the design of effects monitoring programs applicable to the Beaufort Sea. Participants of the workshop are listed in Appendix A. Results of the workshop are incorporated into the remainder of the report.

PRELIMINARY CONSIDERATIONS RELATED TO MONITORING PROGRAM DESIGN

Definition of Environmental Effects Monitoring

For the purposes of this report, environmental effects monitoring is defined as the repetitive measurement of variables to detect changes directly or indirectly attributable to a specific development activity, or to test predicted environmental impacts. It is important to recognize that a definitive demonstration of "no impact" is an impossible goal.

Definition of Long-Term Effect

For the purposes of this report, long-term effects are those which either result from activities which extend over long periods of time or persist as a result of brief activities. Operationally, long-term will be considered to include time periods greater than two years.

ENVIRONMENTAL MANAGEMENT OBJECTIVES

In general, the primary environmental management objective will be to ensure that the effects associated with offshore hydrocarbon development in the Beaufort Sea do not exceed "acceptable levels". At the present time, this objective is extremely difficult for environmental managers to meet because, except for rare exceptions, there is essentially no agreement among policy makers, managers and scientists as to the level of environmental change that is considered to delineate the boundary between what effect is acceptable and what effect is unacceptable in offshore Beaufort Sea hydrocarbon development. The concept of a maximum acceptable effects level is fundamental to the design of effects monitoring programs because it affects the number of samples (and hence cost) required to resolve the effect from the natural variability. In the absence of specified levels, scientists must use their best judgement to set meaningful interim levels for use in the design of effects monitoring programs.

It is a high priority that government and industry should reach an *a priori* agreement on maximum acceptable effects levels of environmental disturbances. Such levels should be set given an understanding and acknowledgement that (1) no development activity in the offshore Beaufort is without concomitant environmental effects; (2) some degree of effect is acceptable; and (3) no monitoring program can ever show that there is zero effect on a particular resource due to offshore hydrocarbon development. Furthermore, government and industry should jointly ensure that a monitoring program is adequate to evaluate whether the levels of disturbance (effect) are being exceeded. Once maximum acceptable effects levels are specified, managers can recommend or impose changes to operating practice or equipment design, if these levels are exceeded. An integral part of the complete process is to involve participation by local native groups, particularly when defining the maximum acceptable effects levels.

REASONS FOR MONITORING

The primary reason for monitoring is to address the uncertainty associated with environmental impact predictions. Two somewhat different types of monitoring arise from this reason:

1) Monitoring to determine whether a predicted detrimental effect occurs, and that the direction and magnitude of the effect is about the same as predicted; and

2) Monitoring to provide assurance that harmful effects that were predicted to be insignificant are, in fact, insignificant.

The latter kind of monitoring has been termed "comfort" monitoring. A secondary reason for monitoring that may arise during the life of a development is to assess the effectiveness of mitigative measures.

Two essential types of information are required for the satisfactory resolution of cause-effect relationships. First, the cause-effect relationship being monitored must be understood to the point that appropriate parameters can be chosen to produce meaningful test criteria. Second, each factor and its relative contribution to the natural variability of the index parameters must be known to the degree required by the pre-set level of confidence expected in the monitoring result. When the above information is unavailable, monitoring cannot proceed. In its place, basic research programs must be developed to provide the essential information.

Environmental monitoring is intended to be operative throughout the life of a project, and should influence the operation of an industrial facility. The process whereby operational protocols are influenced by the results of environmental monitoring programs should form part of a total environmental management approach. It is anticipated, however, that the level of monitoring effort will decrease with time into a development project, as cause-effect hypotheses are tested and the logical conclusion of monitoring is achieved. The results of scientifically valid, project-specific monitoring studies will probably be very useful in either justifying or rejecting the need for monitoring other similar projects.

SCOPE

The emphasis of this study is on producing a framework for designing effects monitoring programs in the Beaufort Sea. This framework should have the following attributes:

1. it should be based on general principles rather than specific detailed steps. This recognizes that, in the end, no two effects monitoring studies will be the same;

2. it should stress the generic aspects of effects monitoring;

3. it should focus on measuring change in the chosen monitoring variables on the following bases:

(i)	temporal -	short-term vs. long-term
(ii)	spatial (1) -	based on natural oceanographic boundaries i.e., nearshore vs. offshore
	spatial (2) -	based on scale of effects; i.e., local (near-field) vs. regional (far-field)

4. it should focus on routine aspects of operations, not catastrophic events such as oil spills;

5. it should exclude compliance monitoring;

6. it should recognize the unique constraints placed upon effects monitoring studies by the arctic marine environment and the subsequent need to stress practical approaches to study designs (measurements);

7. it should be simple and straightforward;

and 8. it should provide guidance to industry.

The intended purpose of this study, then, is not to provide a "cookbook" for designing environmental programs in the Beaufort Sea. Rather, it is to describe the key elements in a logical reasoning process, illustrated with both generic and specific examples, and to highlight important facts or acquired experience that will aid in producing more focused, more cost-effective and more scientifically defensible Beaufort Sea effects-monitoring studies than has been the case in the past.

ELEMENTS OF THE BEAUFORT SEA EFFECTS MONITORING PLAN

There are five, somewhat inter-related elements of the Beaufort Sea Effects Monitoring Plan:

1) planning/experimental design of the monitoring program;

2) field sampling and measurements;

3) sample/data analysis and interpretation;

4) reporting and data management; and

5) quality assurance/quality control (an integral part of the other four elements).

As a general rule, the approximate level of effort (measured as allocation of time or proportion of budget) directed at each element of the plan would be 10-20%, 30-40%, 10-40%, 5-10% and 10-20% for elements 1, 2, 3, 4 and 5, respectively. The remainder of this report deals primarily with elements (1) and (4).

The various components of the planning/experimental design phase are shown schematically in Figure 1. This phase comprises six major stages, most of which contain a number of components. The six major stages are:

1. identification of activity/receiving environment interactions;
2. quantification of the scale/extent of disturbance;
3. assessment of possible effects;
4. formulation of null impact hypotheses;
5. design of the monitoring program; and
6. execution of the monitoring program.

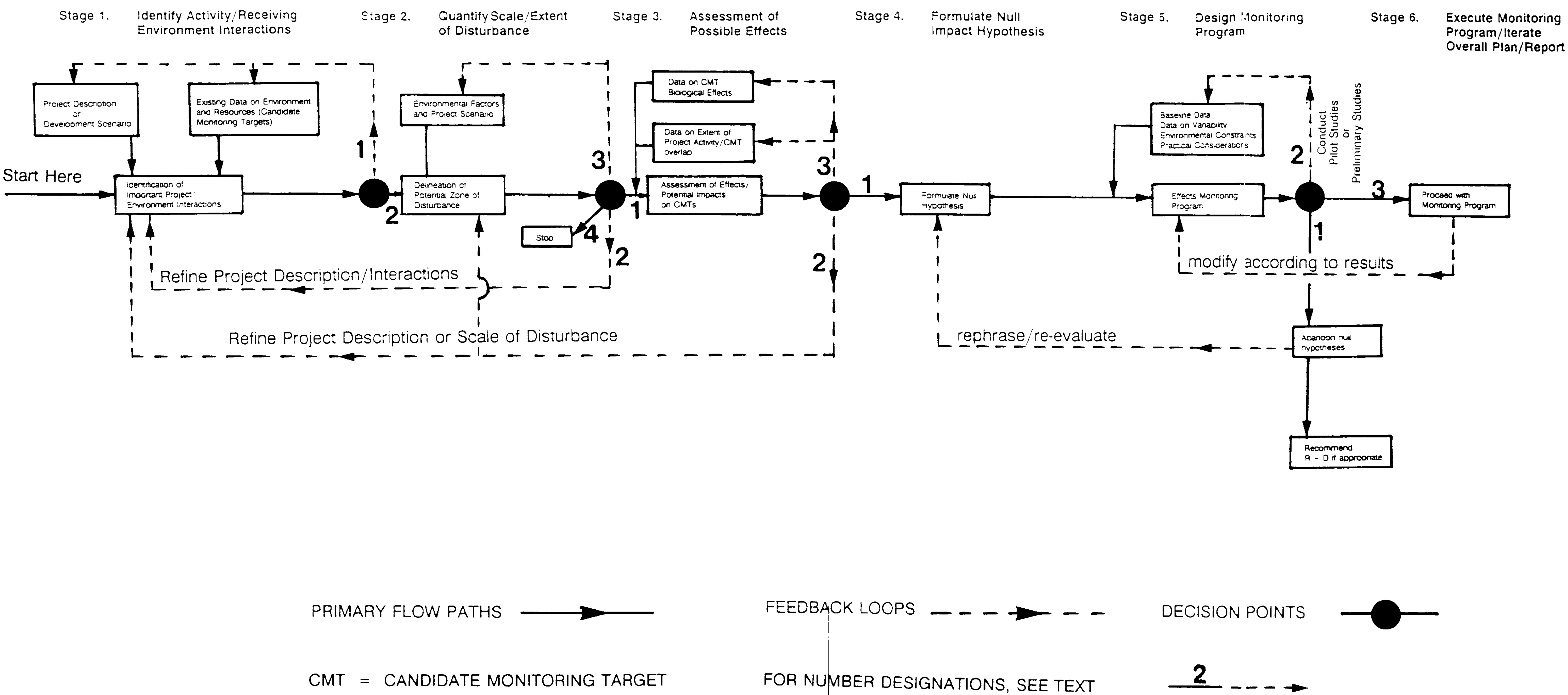


FIGURE 1. COMPONENTS OF THE PLANNING / EXPERIMENTAL DESIGN PHASE OF THE BEAUFORT SEA EFFECTS MONITORING PLAN

The various stages are linked together by feedback loops which permit iteration of the procedures in individual stages of the plan. The procedures and components associated with each stage of the plan are discussed in the following sections. During the planning process, analytical rigour must be applied. So too must be a good measure of healthy skepticism.

STAGE 1: IDENTIFICATION OF ACTIVITY/RECEIVING ENVIRONMENT INTERACTIONS

Stage 1 of the planning process is to identify potential interactions between the offshore hydrocarbon development activities and the receiving environment. To accomplish this, two types of information are required:

(1) a detailed project description or development scenario which relates the types and timing of activities in the Beaufort Sea including utilization of habitats, transportation corridors, construction activities, operational procedures, waste discharge patterns etc.; and

(2) existing data on the state of the receiving environment which could overlap with the activities, including biological resources, environmental quality and the spatial and temporal variations inherent in each.

A convenient way to organize the information is in an "interaction matrix" with activities on one axis and environmental receptor on the other. An example is shown in Table 1. The environmental receptor is generally an environmental resource of value to humans. These resources include those of direct economic value, such as fisheries or waterfowl harvest, but also include those which may be of little or no economic value, but have intrinsic value to society. Examples of the latter include endangered species, marine mammals and rare or aesthetically pleasing environments. In addition to direct

TABLE 1. An example of an interaction matrix relating activities/disturbances associated with offshore petroleum development in the Beaufort sea and CMTS. Each "X" indicates that an interaction is possible.

DISTURBANCE/ACTIVITY	Marine Mammals					Birds						Fish				Lower Trophic Levels				Environmental Quality	
	White Whale	Bowhead Whale	Ringed Seal	Bearded Seal	Polar Bear	Loons	Swans	Geese	Ducks	Shorebirds	Gulls	Pelagic Marine	Demersal Marine	Anadromous	Freshwater	Epontic	Plankton	Epibenthic	Infauna	Water Quality	Air Quality
Human Presence/Shorebases					X	X	X	X	X	X	X										
Solid Wastes					X						X										
Waste Water Discharge			X	X		X			X		X	X	X	X	X	X	X	X	X	X	
Air Emissions					X	X	X	X	X	X	X										X
Routine Oil Spills	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Routine Drilling Discharges	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Airborne Noise	X	X	X	X	X	X	X	X	X	X	X										
Dredging																					
Ice Breaking	X	X	X	X		X			X	X	X	X	X	X			X	X	X	X	
Underwater Sound	X	X	X	X																	

interaction with those resources, we must also consider effects on the marine and coastal ecosystems of the Beaufort Sea Region which support these resources. In this exercise, these important resources of the receiving environment are labelled "candidate monitoring targets" or CMTs. The definition of CMT is broader than that of its BEMP counterpart, VEC, in that it includes such items as water quality or general environmental quality to be used as the focus of monitoring studies. In BEMP, only VECs could be used as a focus for monitoring, VECs being defined as those species that either (1) are important to human populations; (2) have national or international importance; or (3) provide support for VECs under (1) or (2). The advantage of the CMT is that non-VECs can be included in a monitoring program to provide general information, including spatial and temporal trends, on marine environmental quality. Such information, although not directly linked to a VEC can be of considerable importance to environmental managers as an early warning of environmental degradation.

Criteria for the selection of appropriate CMTs include:

1) the CMT must be known or suspected to be sensitive or vulnerable to a particular disturbance;

2) an abundant (biological) CMT is better than a scarce one;

3) it is preferable for the CMT to be distributed homogeneously rather than heterogeneously throughout the monitoring area.

4) sufficient data should be available to establish temporal and/or spatial natural variability in the CMT; and

5) the CMT should be relatively easy to sample, measure, or observe.

Sources of Information

As development scenarios are always evolving, both minor and major changes can occur at almost any time. It is important, therefore, to use the most up-to-date version possible. These scenarios are generally available from the project proponents.

An excellent source of information on various aspects of the biological resources of the Beaufort Sea is the Arctic Data Inventory and Appraisal series published by Fisheries and Oceans. To date, inventories have been published for seals (Harwood *et al.* 1986), bacteria, plankton and epontic communities (Woods and Smiley 1987), whales (Norton *et al.* 1987), marine zoobenthos (Wainwright *et al.* 1987) and fish (Ratynski *et al.* 1988).

As a result of research funded by the Northern Oil and Gas Action Plan (NOGAP), significant advances have been made in recent years with regards to understanding the processes that control the distribution of biological resources and contaminants on the Beaufort Sea Shelf. The results of the studies are extremely important to the design of monitoring programs in the Beaufort Sea because they provide a deeper understanding of spatial and temporal variability. Existing and anticipated NOGAP reports should be reviewed thoroughly during the design of effects-monitoring studies.

Other significant sources of information related to effects monitoring in the Beaufort Sea include ESRF and PERD (Panel on Energy Research and Development) project reports.

Industry-Environment Interactions

Potential interactions between oil industry operations and support activities and the biophysical environment of the Beaufort Sea are relatively well documented as a result of numerous assessments completed for site-specific and regional development proposals. While a large number of interactions between the industry and the environment are possible (as described in detail in the 1982 Beaufort-Mackenzie Delta EIS, Dome *et al.* 1982), most would not be significant from more than a local perspective or affect the abundance, distribution, reproductive potential, survival or harvest of regional populations of biota. Environmental assessments conducted over the past two decades have identified the following key interactions and areas of concern or uncertainty related to oil and gas exploration and potential production in this region (only those issues relevant to foreseeable hydrocarbon development scenarios and considered valid local or regional concerns are identified):

* effects of underwater noise created by stationary (drilling platform) and mobile (ship traffic, seismic operations) on the distribution and numbers of bowhead whales (BEMP Hypothesis No. 1);

* effects of offshore facilities, icebreaking and open-water ship traffic on the distribution, numbers and harvest of beluga whales (BEMP Hypothesis No. 2);

* effects of various activities and facilities in the offshore on populations of ringed and bearded seals (BEMP Hypothesis No. 3);

* effects of icebreaker traffic in the transition (shear) zone on bearded seal pup production (BEMP Hypothesis No. 5);

* attraction of polar bears to active offshore structures leading to increased mortality because of the need to destroy problem animals (BEMP Hypothesis No. 7);

* effects of chronic (episodic) oil spills resulting from routine industry operations on polar bears (BEMP Hypothesis No. 9);

* effects of chronic (episodic) oil spills resulting from routine industry operations on certain species of birds (BEMP Hypothesis No. 11);

* effects of uptake of hydrocarbons released at shorebases and shallow-water production facilities on fish harvest as a result of tainting (BEMP Hypothesis No. 13A);

* effects of the bioaccumulation of specific heavy metals (mercury and cadmium) on fish harvest and human health (part of BEMP Hypothesis No. 13B);

* effects of nearshore structures (e.g., causeways) on coastal temperature and salinity regimes and possible implications to broad whitefish and Arctic cisco populations (BEMP Hypotheses No. 14 and 15);

* effects of drill cuttings contaminated with oil-based muds (OBM) on the populations and harvest of fish, birds and mammals (BEMP Hypothesis No. 20);

* effects of drilling and support activities on the populations and harvest of ducks, swans and geese;

* effects of drilling and support activities on populations and harvest of whitefish, Pacific herring, and burbot;

* effects of drilling and support activities on polar bear numbers and harvest; and

* the long-term effects of drilling wastes and minor spills on fish, seals, birds and bears in the Beaufort Sea food chain.

As noted above, certain key industry-environment interactions have been identified and are the focus of monitoring studies. Nonetheless, it remains prudent to begin new monitoring designs by compiling an up-to-date interaction matrix (similar to that shown in Table 1) particularly when new development scenarios are proposed, when new locations or times for activities occur or when the focus of the monitoring is site specific or in the shallow nearshore zone.

At the completion of Stage 1, one of two decisions must be made:

1) if it is judged that the identification of important project/environmental interactions is inadequate or not complete then feedback within Stage 1 to obtain more or better information on the project description or existing environmental data; or

2) proceed to Stage 2.

STAGE 2: QUANTIFICATION OF THE SCALE OF THE POTENTIAL DISTURBANCE

A key step in the scientific design process is to make some estimate of the scale (extent and duration) of the disturbance being evaluated. The scale can be expressed as a "zone of influence", a "zone of potential effect", or a "zone of disturbance" of short-term (temporary) or long-term (continuous or persistent) duration. The concept applies equally well to all types of disturbance. An example, involving contaminants, is used below to illustrate how the estimation of a zone of disturbance is used in the design process.

Suppose that a concern arises over the possible discharge of chromium during drilling operations and the possible bioaccumulation of chromium into marine biota. Clearly, if one litre of drilling waste containing 1 ppm of dissolved chromium was discharged to the sea, no rationale for a monitoring program could be supported. Conversely, if drilling effluent containing 1% chromium was discharged continuously to the sea at a rate of 1 $m^3 s^{-1}$, a monitoring program would be imperative and a high priority. What is less clear and often a judgement call is where the boundary exists between needing to design or not needing to design a program. A mass balance calculation can help in many cases, particularly when the decision is based on physical or chemical arguments. In this example, assume that the discharge is 10 litres per minute (continuous) and that the dissolved chromium concentration in the effluent is 10 ppb. As a first approximation assume that the initial dilution is zero. At the specified rate of discharge, one day's discharge would result in a 14.4 m^3 slug of water containing 10 ppb chromium. Offshore, this would probably be seen as insignificant, and not worthy of additional consideration. In very shallow water, perhaps some reason would exist to consider the matter further, perhaps not. If further consideration in either location was considered prudent, then the fate of the effluent in the receiving environment would be considered.

Specifically, the first approximation of zero dilution is not realistic. We know from many observations in the Beaufort Sea that typical minimum dilutions of 100 fold can be expected to occur during waste discharge (Thomas *et al.* 1983). In the current example, this is intuitively comfortable because it requires only 1 440 m^3 of water to dilute one day's discharge. At 100-fold dilution, the incremental change in chromium concentrations would be 0.1 ppb. Given that average concentrations of chromium in the Beaufort Sea are in the order of 0.5 ±0.4 ppb, and that sampling/analytical variance would be at least ±30%, it is clear that such an increase would not be detectable. There would, therefore, be no point in designing a program because the minimum expected change would go undetected. When the dispersive effects of currents, tides and density mixing are considered, then the degree of dilution would only increase, indicating that indeed a very conservative approach has been taken.

The key to using Stage 2 effectively is to have quality information about the concentration of chromium in drilling effluents, the concentration of chromium in receiving waters and the nature of currents, tides and other physical factors in the receiving environment. Such information is available from several sources. The chemical characteristics of routine waste discharges in Beaufort Sea drilling are summarized in INAC 1986a, INAC 1986b, INAC 1985, and Thomas *et al.* 1983. Information about contaminants in the Beaufort Sea receiving environment are given in Thomas *et al.* (1982), while information on the physical environment are catalogued in Birch *et al.* (1982).

In situations where site-specific information is not available, the mass balance approach can still be applied by extrapolating data from nearby areas or similar drilling circumstances and making the appropriate allowance for the lack of specific information. In some situations, the issue under consideration may be of such importance that specific information must be obtained. In such cases, feedback to Stage 1 may be necessary to obtain clarification or new information.

At the completion of Stage 2, one of four decisions must be made:

(1) proceed to Stage 3;

(2) go back to Stage 1 for further information;

(3) iterate within Stage 2 for more information;

or (4) stop at Stage 2 because the project environment/interaction does not warrant further consideration on practical or mass -balance grounds. This must only be done by considering the resources of concern (CMTs). For example, a possible effect which occurs over 1 km^2 of rare, critical or exceptionally valuable habitat and occurs during a relatively short, but critical time period may be of greater concern than one which occurs over 100 km^2 of more widespread habitat for a relatively long time.

STAGE 3: ASSESSMENT OF POSSIBLE EFFECTS

Stage 3 is a focusing step in the design process. In this stage a great deal of information linking physical, chemical and biological systems and the results of Stages 1 and 2 is distilled into an assessment of an actual zone of effect as compared to the potential zone of effect of Stage 2. Scientific judgement and experience and knowledge of the Beaufort Sea region are essential requirements to complete this stage properly. The result of Stage 3 is important, for it leads directly to the formulation of hypotheses which provide the specific focus for any monitoring program.

Stage 3 is directed primarily at assessing the impacts of industry activity on biota in the receiving environment. The assessment must address the following questions:

(1) how probable will the impact be?
(2) how persistent will the impact be?
(3) how pervasive will the impact be? and
(4) how significant will the impact be?

A convenient process for answering these questions is described in Duval *et al.* (1990). Briefly, the assessment process would comprise the following:

1. composition of an interaction matrix which identifies possible interactions between project disturbances and valued environmental components (similar to Stage 1 in the present process);

2. a semi-quantative determination of the extent of spatial and temporal overlap between project disturbances and the valued environmental components using the concept of the "zone of influence" of each project disturbance; and

3. determination of environmental significance by testing the interaction with the following questions:

    (i) is the effect associated with the project disturbance likely to change the reproduction or survival of individuals in the valued component(s) or the productive capacity of habitat?
    If Yes, then proceed to Question (ii)
    If No, then the interaction is CLASS 4 INSIGNIFICANT

    (ii) given the spatial and temporal overlap, would the effect on the valued components or habitat be of concern?
    if YES, then proceed to Question (iii)
    if NO, then the interaction is CLASS 4 INSIGNIFICANT

(iii) would the anticipated effects on habitat productive capacity, survival or reproduction of the population be considered unacceptable?
if YES, then proceed to Question (iv)
if NO, then the interaction is CLASS 4 INSIGNIFICANT

(iv) will recovery of the population or habitat productive capacity occur?
if YES, then proceed to Question (v)
if NO, then the interaction is CLASS 1 SIGNIFICANT

(v) how long is the recovery expected to take?
if long-term, then the interaction is CLASS 2 INSIGNIFICANT
if short-term, then the interaction is CLASS 3 INSIGNIFICANT

This approach is not the only valid way of assessing environmental significance. Others may be just as effective. However, the approach described above has a very attractive feature with regards to the design of monitoring programs. Specifically, it incorporates a permanent record ("audit trail") of assumptions, calculations, evidence and justifications for all decisions made throughout the assessment process including the perceived quality of the information being used. This provides a measure of quality control and allows incomplete or inadequate data to be identified. In terms of monitoring program design, it would help identify (1) information needs critical to program design; (2) relevant research needs. In addition, it provides a systematic basis against which improvements to the monitoring program can be made as monitoring results become available.

In certain situations, an assessment of possible effects of a given development scenario or site specific project may be completed by groups or individuals outside the immediate monitoring design process. Direct use of such results should be strongly resisted if these results are used in place of completing stages 1 and 2 of the monitoring design process. This is because the full value of a carefully thought out and effective monitoring program may be lost if there is an interruption in the logic and consideration of subtle, but important factors as would probably occur if the design team scientists did not remain fixed on a specific monitoring objective from the beginning of the planning process to the end.

The completion of Stage 3 leads to another decision point. There are three possibilities:

(1) proceed to the Stage 4 (formulate null hypotheses);

(2) if the assessment is considered inadequate or unacceptable, attempt to rectify the problem by acquiring additional information and repeating Stage 3; or

(3) as for (2) above, but loop back to Stage 1 or Stage 2.

STAGE 4: FORMULATION OF A NULL IMPACT HYPOTHESIS

A critical element in the design of an environmental effects monitoring program is to clearly establish exactly what the objectives of the program are. Ideally, the objectives should have clear relevance to the environment and to managers who have the responsibility of managing environmental protection. Furthermore, the objectives must define the specific effect to be monitored and the level of effect that identifies a pass/fail boundary, for example, a specific concentration level of hydrocarbons in sediments.

An appropriate means of organising an effects monitoring program is the testing of null hypotheses. Null hypotheses have three essential elements:

(1) they must specify the spatial scale on which it is desired to observe differences in the monitored variable;

(2) they must specify the temporal scale on which it is desired to observe differences in the monitored variable; and

(3) they must specify the resolution of measurements of the monitored variable; i.e., the magnitude of the smallest change that the program must detect at a specified level of statistical significance.

While few previously conducted effects monitoring programs in the Beaufort Sea have dealt adequately with spatial and temporal scales, virtually none have incorporated the required resolution into the planning phase of a program. Rather, resolution has usually been ignored until sampling and analysis are complete. This has often resulted in a waste of resources because either (1) the sampling program was inadequate and failed to provide the required resolution (no change observed or verified because minimum observable change was too high); or (2) the sampling program was intensive and identified a change so small as to be of no environmental significance. Establishment of the required resolution can be an involved process as it must be (1) scientifically attainable; (2) attainable through a sampling and analysis program which can be accomplished with available resources or logistical constraints; and (3) environmentally significant. The selected resolution of measured changes or differences is not environmentally significant if it is (1) so small that there is no potential threat of environmental damage; (2) so large that there is no credible scenario in which it could occur; or (3) so large that, once detected, substantial environmental damage would already have occurred.

Selection of an appropriate pass/fail threshold and resolution of measured change requires specific information on the effects of the various anthropogenic activities or inputs, the manner in which the receiving environment responds to stresses of different types and magnitudes, the structure and variability of biological populations and the physical and chemical attributes of the receiving environment.

Null hypotheses take the general form:

H_O : An activity has not caused a specific attribute of the receiving environment to change by a specified amount (magnitude [resolution], geographical and temporal extent of change must be specified).

As many attributes of the receiving environment may also be affected by natural changes, the above hypothesis must be restated as two dependent hypotheses of the form:

H_O (1) : A specific attribute of the receiving environment has not changed by a specified amount (or does not differ from defined background conditions)

H_O (2) : Changes in the specific attribute in H_O (1) by the specified amount are not related to the activity.

A specific example:
The overall null hypothesis is:

H_O : Discharge of wastes during offshore exploratory drilling at the Adgo location will not cause the concentration of chromium in sediments to increase above background levels beyond 100 metres of the discharge location.

The related dependent hypotheses are --

H_O (1) : The concentration of chromium in sediments have not increased beyond background levels within 100 metres of the Adgo location.

H_O (2) : The changes in chromium concentrations within 100 metres of the Adgo location are not related to offshore oil and gas exploration activities in the Adgo area.

The second (dependent) hypothesis need not be tested and is, in fact, untestable unless the first is tested and disproven. As the establishment of cause/effect relationships is often very difficult, monitoring programs which address H_O (1) alone are much simpler and more likely to succeed. In those instances where change beyond the preset levels is detected, specific research or monitoring studies to address H_O (2) can be performed. Where observed changes are less than the preset levels, further studies fall clearly into the realm of research and do not belong as part of a monitoring program.

During the program design phase considerations of various null hypotheses may focus the monitoring program by deleting or altering some of the null hypotheses. For example, extreme spatial or temporal variability may preclude the achievement of an environmentally significant resolution. Resource limitations may prohibit sufficiently intensive sampling and analysis. In such cases, testing of the hypothesis should not be performed, because it can only result in equivocal results. An alternative null hypothesis directed at some other indicator parameter or effect would be advised.

In other cases, it may be possible to show, by calculation, that the maximum potential change that could occur in a monitored parameter is so small that there is essentially no finite probability that it could exceed the pass/fail threshold. In these cases, the null hypothesis is unsuitable and should be eliminated from the monitoring program. On the positive side, a properly performed design study will lead to a less expensive monitoring program wherein the desired resolution will be successfully achieved through a more limited effort. Once Stage 4 is complete, the remainder of the design phase should be completed by considering each null impact hypothesis separately.

STAGE 5: DESIGN MONITORING PROGRAM

A monitoring program is a measurement experiment in which the only variables under the control of the experimenter are: (1) where and when the samples are taken; (2) how many samples are taken; and (3) what is measured. Designing effective monitoring programmes entails making appropriate decisions with regard to objective, statistical model, sampling design and methods of collecting and analysing samples. These decisions should not be made independently. Before the required number of samples can be calculated, the desired sensitivity for detecting impacts must be decided and specified in terms of a statistical model. The statistical model defines the hypothesized nature of the CMTs to be sampled and is based on a knowledge of the physical, chemical and biological nature of the system and the study objectives (e.g., just to detect an effect or also to delimit spatial extent). The sampling design follows from the statistical model, specifying exactly how, where and when or how often the samples are to be collected as well as how many and what parameters are measured.

Hypothesis testing will usually involve a comparison of population means as a way of deciding whether an effect has occurred. The simplest kinds of statistical models to accomplish this one are the one sample **t**-test and the two sample **t**-test. The former enables one to test the hypothesis that a sample was drawn from a population which has a true mean less than or equal to a specified value. The one sample **t**-test would be applicable, for example, in a situation where a value of a chemical contaminant was compared to a fixed regulatory limit. The two sample **t**-test enables a test of the hypothesis that there is no decrease in beluga whales relative to that which existed prior to a development event. More complicated models are required to test hypotheses involving pre- and post- event samples take at impact and reference locals. This involves comparison of 4 population means and is called the optimal impact assessment design. In this case an impact is defined as a change in the CMT that occurs in the impact area, but not in the reference area.

The need for more complex models would result if the impact area might be stratified or if a suite of parameters is used in the hypothesis (multivariate model).

Design of the monitoring program will involve making decisions on the following:

(i) what to sample,

(ii) sample replication (sample size),

(iii) where to sample,

(iv) when to sample, and

(v) how to sample (methods).

Guidance for making decisions on the above will include baseline data on biological, chemical and physical parameters, experience in previous studies in the area, and scientific judgement.

What to Sample

Once the appropriate CMT has been identified for a particular hypothesis, a decision must be made as to which parameter or suite of parameters can be used most effectively in testing the null hypothesis. The choice will depend on cost, relevance and knowledge of the successfulness of using the parameter(s) in previous monitoring programs related to the effects of offshore hydrocarbon development both in the Beaufort Sea and in other areas.

In some cases, the choice of parameters will be straightforward; for example, beluga whale distribution in evaluating the relationship between noise and beluga whale distribution. In other cases, a large number of parameters will be possible. Such is the case in monitoring the effect of contaminants on biota, such as fish or benthic invertebrates. Not all biological variables will be equally appropriate for monitoring contaminants on these biota, and their suitability for inclusion may be addressed in terms of a number of criteria as listed below (adapted from IMCO *et al.* 1980).

Ecological Criteria

(i) ecological significance: can the effect be shown, or convincingly argued to be related to an adverse or damaging effect on the growth, reproduction or survival of the individual or the population and ultimately on the well-being of the community/ecosystem?

(ii) relevance to other effects: can the effect be related to other effects at higher or lower levels of organisation?

(iii) specificity: how specific is the effect in relation to the causative agent?

(iv) reversibility: to what degree can the variable return to its original level when the causative agent is removed?

(v) range of taxa: is the effect specific to particular taxa?

Efficiency Criteria

(i) quantitative aspects: does the effect bear a quantitative or predictable relationship to the cause?

(ii) sensitivity: what intensity of stressor is required to elicit the response?

(iii) scope: over what range of intensity of stressor is the effect observable?

(iv) response rate: how quickly is there an observable effect? hours? days? years?

(v) signal to noise: can the effect (signal) be easily detected above the natural variability (noise)?

(vi) precision and accuracy: can the effect be measured accurately and precisely?

Administrative Criteria

(i) cost: how expensive is the measurement of the variable in terms of capital equipment, operational costs (manpower)?

(ii) application: to what extent has the effect been used in a field monitoring program and shown to be related to the cause (in this case contaminants)?

(iii) relevance to Beaufort Sea: is there anything about the use of the measurement that would be precluded by the isolation and special operating constraints occurring in the Beaufort Sea?

Using the above criteria a number of biological variables were evaluated for potential usefulness in contaminant effect monitoring studies in the Beaufort Sea (Table 2). The ratings were categorized as follows:

TABLE 2

Evaluation of suitability of various biological variables for monitoring the effects of contaminants in the Beaufort Sea.

	Measurement	Evaluation
1.	Ecological Effects	
	Community biomass	+++
	Abundance	+++
	Diversity indice	+++
	Changes in size class distribution of populations	+++
	Species density	+++
	Growth rate	+++
	Reproduction	+++
	Population Structure	+++
2.	Biochemical Effects	
	specific	
	mixed function oxidase	++
	metallothionein	++
	bioaccumulation	+++
	non-specific	
	lysosomal stability	++
	steroids	+
	blood chemistry	+
	adenylate energy change	+
	taurine/glycine	+
	primary production	++
3.	Physiological Effects	
	Respiration	+
	Feeding Rate	++
	Body condition indices	+++
	Scope for growth	++
	Oxygen (consumed)/Nitrogen (excreted)	++
4.	Morphological and Pathological Effects	
	neoplasia/tumours	+
	liver somatic index	+++
	liver structure	++
	ulcers	+++
	fin erosion	+++
	gill deformity	++
	gametogenic cycle	+
5.	Genetic Effects	
	chromosomal abnormalities	+
	mutagenicity assay	+
6.	Behaviourial Effects	+
7.	Bioassay	++

+++	highly recommended and appropriate for use in Beaufort Sea on a routine basis
++	recommended for selective use only because of lack of proven record or limited by Beaufort Sea location
+	not recommended because of unproven relevance or because impractical in frontier setting of the Beaufort Sea

Careful consideration of appropriate monitoring parameters may eliminate from the monitoring program some of the parameters that have been routinely measured as a matter of course. Although there may be some reluctance in dropping traditionally measured parameters, it may be necessary because the monitoring program will be weakened by including parameters that cannot provide unequivocal data relevant to testing of the impact hypothesis. It is also important to note that many parameters traditionally monitored are indicative of the same effect. For example, various sediment chemistry measurements, benthic infaunal population studies and an array of benthic fauna biochemical stress measurements are all used to monitor damage by pollutants to benthic communities. It is neither efficient nor necessary to monitor all these parameters if the evaluation of harm to the benthic community can be unequivocally made by measuring just one or two of the parameters in a statistically sound monitoring program.

Another consideration in what to sample involves choice of supporting measurements which are essential to final interpretation of results. For example, during aerial observations of whale distribution it is usually necessary to take note of the position of the Mackenzie River sediment plume or the amount and type of ice present. Similarly, collection of sediments for chemical analysis is incomplete if samples are not also collected for grain size measurements. The number and type of measurements or observations on samples to be taken in support of the primarily study objective will depend on the goal of the monitoring program, and the level of scientific understanding of the relationship between primary and supporting data. The list of supporting measurements can be extensive, as illustrated by contaminant levels in organisms where species, weight, age, sex, lipid content, season, reproductive state, salinity, temperature and other interactions are all important factors in interpreting data (Phillips 1980).

During the workshop, specific comments and guidelines related to fish, mammals, birds, benthos and sediments as CMTs as well as practical suggestions for carrying out sampling programs were expressed. These are summarized below.

FISH

- generally no standard methods
- fish not appropriate CMT for offshore studies because marine species such as flounder and arctic cod are too mobile and move into nearshore areas. Natural variability not understood for any offshore species. May also be abundance problems.
- for nearshore studies, anadromous fish are not appropriate CMT because even large (±40%) changes in population would probably go undetected due to lack of adequate stock abundance data.
- due to general lack of baseline data changes in fish population resulting from a major oil spill would probably not be resolved from the natural variability.
- comprehensive data on biological characteristics of fish stock should be obtained before fish considered for offshore, or regional scale effects monitoring.
- fish may be appropriate as CMT in site-specific nearshore areas in conjunction with bioaccumulation of contaminants and possible tainting issues. Relatively stationary species that are found in close association with sediments would be best approach

MAMMALS

- few standard methods for studying bowhead whales, white whales, polar bears or seals. Some do exist and have been used successfully in the area of aerial behaviour observations of bowhead whales, but not of white whales.
- data for population estimates of whales is presently inadequate. Point Barrow estimate of bowhead whale population is at least twice that of estimate (by aerial

surveillance) in Beaufort Sea. The only estimate of white whale population is for Mackenzie River estuary.

- adequate data on how white whales use the Mackenzie Delta area
- whale variables that are considered well enough understood to be amendable to monitoring include distribution, and standard behaviour of bowhead whales at the sea surface.
- monitoring for change at the population level of marine mammals is not recommended because the understanding of population dynamics and natural variability is inadequate. It would be more appropriate to monitor reproduction rates of some mammals, such as seals, where changes can be detected more quickly and more reliably.
- Marine mammals are good CMTs for contaminant monitoring. Studies are already underway involving uptake of LRTAP-related contaminants, particularly organochlorines in white whales, seals and polar bears of the Beaufort Sea region.

BIRDS

- no standard methods used
- generally not appropriate CMT in offshore because most offshore birds are migratory so that they can be exposed to other impact sources in areas outside the Beaufort Sea.
- effects monitoring studies involving birds may be very important in coastal breeding areas particularly where these areas may overlap gas developments in the Mackenzie Delta. Effects monitoring would be directed at ensuring that mitigative measures such as flight corridors are successful in limiting impacts.
- loons would be a good CMT for monitoring in coastal areas.

BENTHOS

- attractive as a CMT because
    - standard methods are available for sampling, analysis and statistical interpretation

- they are abundant
- they are relatively easy to sample
- they are immobile
- can choose species that are always present at a location
- can always find sensitive and non-sensitive species to almost any disturbance
- can detect changes in benthic communities with relative ease
- provide means of long-term monitoring
- good for monitoring "surprises"
- considerable database exists for effects of offshore petroleum development on benthos, for example, Appelbee and Mair (1981), Davies *et al.* (1984), Addy *et al.* 1984, Kingston (1987), NRC (1983), Hargrave and Thiel (1983), Harper *et al.* (1981), Battelle/Woods Hole (1983), Reiersen *et al.* (1989).

- no baseline for much of the Beaufort Sea is a problem, but can still be used effectively to monitor a gradient of effect
- benthic studies most appropriate in water depths greater than 2 m to avoid areas of inherently huge natural variability caused by suspension/sedimentation and ice scour.
- pilot and baseline studies essential part of monitoring plan for benthos studies to establish natural variability on time and space scales appropriate to the hypothesis being tested (small scale variability can be very large).
- little value in near-field or near-term monitoring of benthos. Such effects are well-known and well documented. Of greater interest are possible regional and long-term changes to benthic community parameters due to chemical effects or alterations in the physical environment
- need to identify benthos sampling stations during exploration drilling whenever possible to provide data for monitoring studies during the development phase. These controls would be the beginning of a valuable time series and provide needed information on temporal variability.

SEDIMENTS

- are best CMT for studies related to the presence/absence, temporal trends, loading rates and point sources of contaminants
- good database exists for most contaminants in nearshore and offshore Beaufort Sea
- processes controlling contaminant concentrations and processes involving role of sediments in the fate of contaminants relatively well understood
- temporal controls generally unnecessary. If pre-development baselines required a convenient surrogate is the use of historical sediment cores
- good standard methods for collection, analysis and statistics
- easy to sample; samples easily replicated
- sediments are an accumulating medium; consequently sediments integrate contaminant effects
- sediments can be "early warning system" of environmental degradation in that build up of contaminants can often be detected before onset of biological effects
- natural variability of sediment contaminant concentrations on short time and distance scales tends to be small and differences are largely explained by the grain size effect. Gathering of synoptic data, therefore, is straightforward and concentration gradients are usually easily identified
- relatively inexpensive

General Considerations for Field Operations

- build in adequate contingency time for sampling to account for weather delays, "hurry-up-and-wait" delays and unforeseen problems characteristic of arctic field operations.

- bring duplicates of important equipment, triplicates of critical equipment to the field as back-ups

- do not underestimate value of "arctic experience" when planning scientific roster for field operations

- timing for initiation of field operations must be flexible; spring, for example,is not a date on the calendar, rather it is tied to physical events such as break-up of the Mackenzie River, the timing of which may vary from year to year

- use "tried and true" techniques to the extent possible if they are adequate

- make sure equipment can be transported by available aircraft, ship, boat, etc.

- examine sampling locations for safety of operations vis-à-via polar bears, shear ice zone, etc.

- sampling design with birds, mammals must not itself be a significant disturbance

- ice environment of the Beaufort Sea is an over-riding factor in the region; annual variation in the biota are often directly linked to, and can often be explained by, annual variation in the ice environment

Sample Replication

A common problem in planning a survey or monitoring program is that of calculating the degree of sample replication, or sample size **n**, required to obtain an estimate of a specified precision or to test a hypothesis about a difference between two sample means. For example, if the objective is to detect a specified difference between two means with a specified degree of confidence, the formula for calculating the required number of samples is:

$$\mathbf{n} = \frac{2\,(Z_{\alpha}+Z_{1-\beta})^{2}\sigma^{2}}{\delta^{2}}$$

where δ is the magnitude of the effect; i.e., the true difference between the two population means being tested, α is the significance level of the test and ß is the power of the test. An estimate of σ^2 will usually be available from baseline studies, measured in a pilot sampling programme, or else guessed from other prior experience, but values for α, ß and δ will have to be chosen before **n** can be calculated.

Hypothesis testing leads to four possible outcomes as shown below:

TRUTH	CONCLUSION	
	NO EFFECT	EFFECT
No Effect	No Error (prob. = 1 - α)	Type I - Error (prob. = α)
Effect	Type II Error (prob. = ß)	No Error (prob. = 1-ß)

The incorrect rejection of a true null hypothesis is referred to as a Type I error. The probability of a Type I error is designated by α and is referred to as the significance level of the statistical test. The risk of Type I error is borne by the project proponent. This type of error could lead to unnecessary restrictions upon hydrocarbon development. The incorrect acceptance of the null hypothesis is referred to as a ß or Type II error where ß represents the probability of the incorrect decision. ß is also referred to as the power of the test, specifically the power to prove that a null hypothesis is false (i.e., detect an effect when one has occurred). The risk of Type II error is borne by the regulator. This type of error would cause impact to go unnoticed and development to proceed without appropriate attempts at mitigation.

The power of the statistical test is determined by five study design parameters: (1) significance level (α) of the test; (2) number of sampling locations; (3) number of replicates; (4) minimum detectable difference (δ) specified for the monitoring variable; and (5) residual error variance (i.e., natural variability within the system). The greatest increase in power is produced by a reduction in the residual term. Although an increase in sample size can improve the power of the test, a far greater increase in power can be achieved by identifying additional sources of variation and removing them from the residual by design particularly when the residual term is large. To be successful in increasing the power of the test, we must understand the causes of natural variation.

Choosing a value for δ

A central problem in developing a meaningful criterion for sample replication in a monitoring programme lies in deciding what magnitude of effect, δ, should be detected with probability ß. As indicated earlier, δ should be sufficiently small that it is well below the threshold at which ecological damage might occur, so that a deleterious effect would not escape notice, but not so low as to be ecologically insignificant. Obviously, a statistically significant effect is not necessarily an ecologically significant one. An arbitrarily small effect can be detected by making the number of samples very large. As an example, if sediment contaminant monitoring is to be used as a screening method to decide whether biological monitoring is needed, δ could be chosen to be a level which would initiate some type of biological monitoring. Alternatively, if it is not possible to decide what value of δ would trigger an action, then δ could be more conservatively chosen to be the smallest change that can be practically detected or it might be set at the threshold at which a change could be considered to be ecologically significant,if such a threshold can be defined. In general, the number of samples should be chosen to err on the side of the environment. If δ is set at the threshold where damage might occur, the power for the test should be very high. On the other hand, it would be wasteful to set δ at a level below one for which any biological effect could occur. Natural variability (σ^2 for the population being sampled) defines a natural scale for the magnitude of δ, because **n**, the required number of samples, is proportional to $(\sigma/\delta)^2$. Thus as δ is reduced below σ, an increasingly large number of samples is required. Because of this relationship, it is not practical to set a goal for δ that is much smaller than natural variability.

A practical approach to deciding how many samples to collect, therefore, would be set to δ equal to the square root of the natural population variance for the Beaufort Sea as a whole. But would this criteria for δ be low enough to ensure that an ecologically adverse effect would not escape detection? It appears reasonable to assume that a change in contaminant concentration which is well within the range of natural variation for the Beaufort Sea as a whole would not be ecologically harmful. The same argument may hold for biological CMTs, although considerable additional thought must be given to the appropriateness os using $\delta = \sigma$ or some other factor, perhaps $\delta = k\sigma$ where k is <1 and varies in accordance with our level of understanding of the resource. It is emphasised here that the purpose of the current exercise is to outline an approach to setting values for δ, α, ß etc., not to make the actual choice for a given situation.

Choosing values for α and ß

The choice of values chosen for α and ß will be based largely on the perceived consequences of making Type I and Type II errors as well as the cost of obtaining the information.

Calculating the sample size.

Once the values have been set for δ, α , ß and σ^2, the value of **n** can be calculated. Then the cost of collecting the data can be estimated and compared with the budget. If it is found that achieving the goal will require more money than is available, a practical decision will have to be made to scale down expectations (adjust the goal) or increase the budget. If quantitative goals for detecting impacts are not formulated, sample size decisions will probably be dominated by cost. Some possible reasons for not formulating a quantitative goal are: no understanding of the statistical problem; no knowledge of σ^2; or no decision as to how large an effect (δ) should be detectable.

The importance of the level of unexplained variance in determining δ is illustrated in Figure 2. As the unexplained variance (expressed as coefficients of variation of 30, 50, 70 and 90%) increases, the minimum detectable differences between sampling stations also increases. In addition, as the level of unexplained variance increases a greater level of sample replication (and cost) is required to detect a given level of difference. For a more detailed description of the cost/benefit considerations in sample replication, refer to Hoff and Thomas (1986).

The results of data analyses performed to date on Beaufort Sea sediment contaminant data sets have a strong influence on the value of **n** for sampling programs. It has been shown, for example, that sediment grain size can explain from 31-91% (most values greater than 85%) of the variance in sediment trace metals and PAH concentrations in Beaufort Sea nearshore, subtidal and embayment data sets (Yunker, 1985; Hoff and Thomas, 1986; Wainwright and Humphrey, 1988). For most metals and PAH, regression on percent clay explains almost all of the environmental variance. What remains is essentially the error of chemical measurement. The relevance to **n**, the sample size, is that if the variance that can be explained by the grain size effect is not removed, then an unnecessarily large value for σ^2 is obtained which similarly inflates **n** and the cost of the program.

Another factor to consider when establishing sample replication, is the idea of oversampling and archiving. Oversampling is relevant to contaminant issues. It is useful in providing additional samples which can be archived; these samples can be used at a later date (1) to check questionable results; (2) to look for compounds not part of the initial analysis program design; (3) to provide sample for intercalibration of results; or (4) to provide sample for possible future method development/intercomparison.

Where to Sample

Sampling locations are influenced by the objectives and goals of the monitoring program, the choice of CMT, the sampling design and sometimes by natural barriers such as the ice transition zone, rubble fields etc.

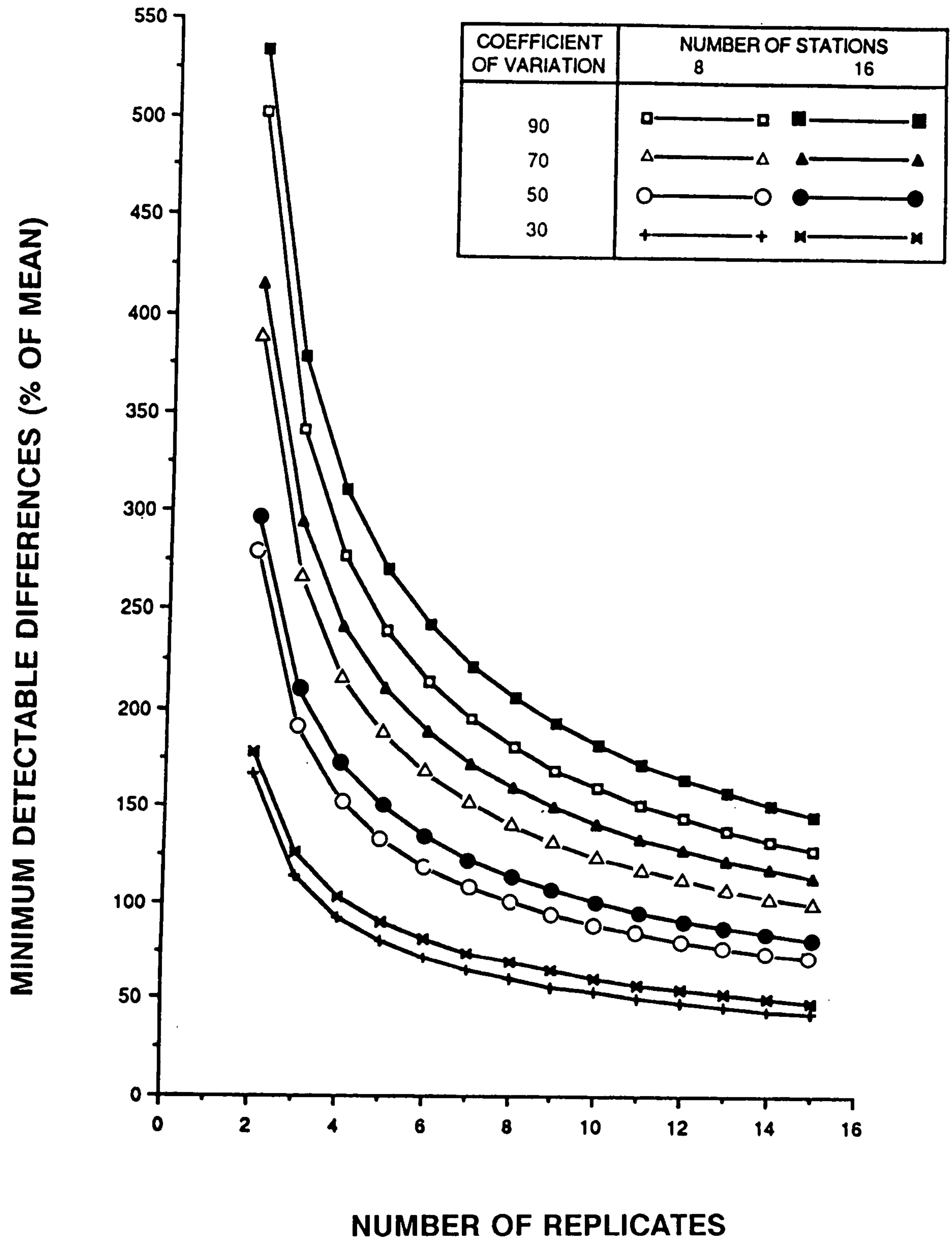


Figure 2. Minimum detectable differences (δ) vs. number of replicates at selected levels of unexplained variance for 8 and 16 stations. β = 0.80, α = 0.05. (after Tetra Tech 1987)

Random sampling should be used when the objective is simply to detect a uniform change in the CMT over a specified area. (Example: contaminants around a point source). Stratified random sampling should be used if a non-uniform distribution of the CMT is anticipated. (Example: benthos around a dredge site). If the objective is to measure the spatial distribution of the CMT within a given area or to define the size of an area which has changed, a systematic arrangement of stations is best (Example: distribution of whales in a drilling vs non-drilling zone). If the goal of the monitoring program is to measure parameters of a dose/response, curve then the sampling locations should be at distances which correspond with equal amounts of change in the response. An example is illustrated in Figure 3. In this case, response decreases exponentially with distance from a source. Consequently, sampling occurs at distances to correspond to equal changes in response (ΔY)

Estimating an area within which sampling should be conducted is related to δ. A straightforward example is provided by waste discharge around a point source. Assuming that we know the loading for contaminant X, that the discharged contaminant settles uniformly around the point source, and that the concentration of contaminant is more concentrated than in the natural sediments, then

$$\text{area} = \frac{\text{Input X (g)} \times 10^{6}\ \mu g.g^{-1}}{\delta\ (\mu g.g^{-1}) \times D\ (g.cm^{-3}) \times 1\ cm}$$

where input X = total estimated input of contaminant X to the sediments

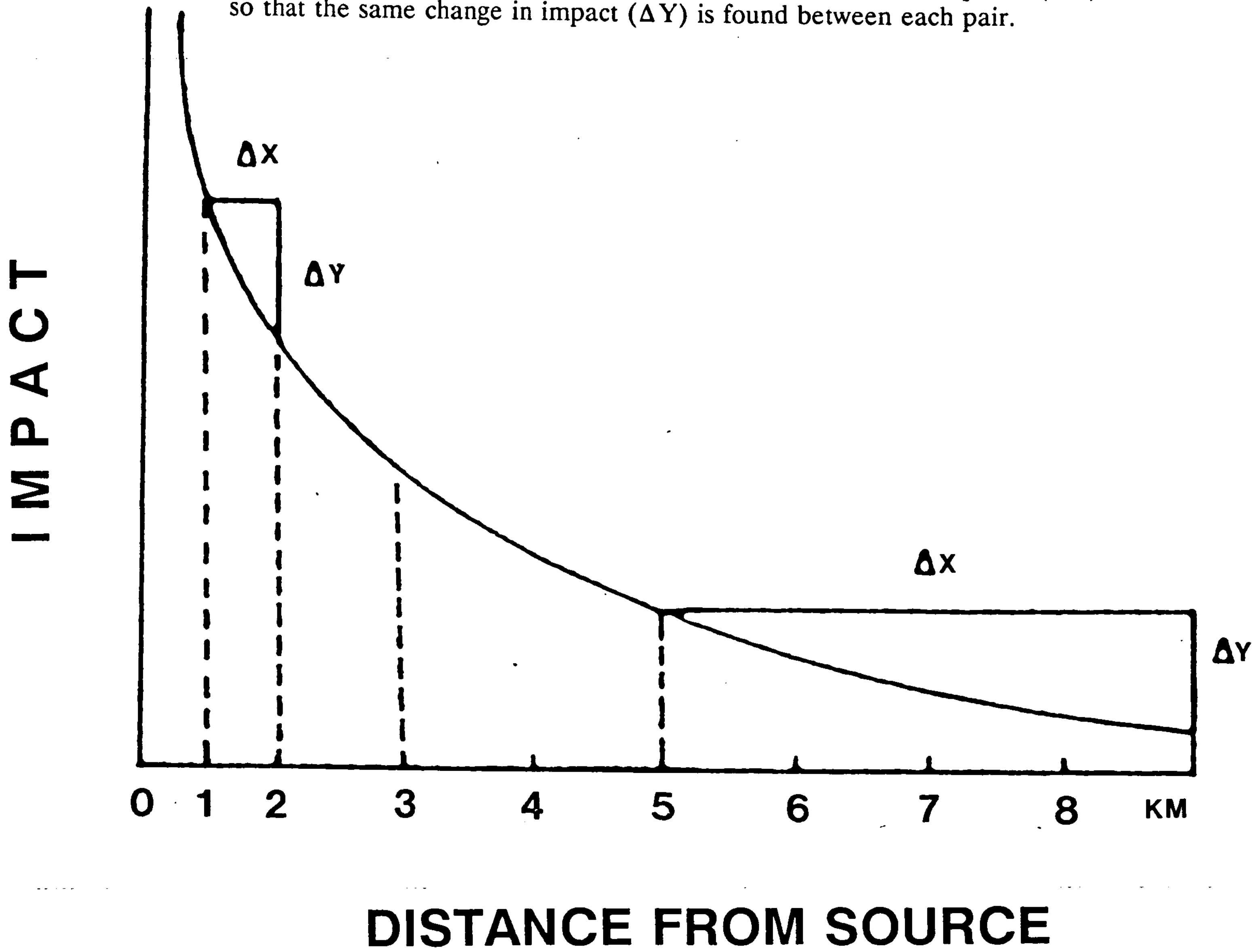
δ = the increase in contaminant concentration that should be detectable

D = density of the sediments

1 cm = upper 1 cm of sediments are sampled

A more complicated formula could be derived to take into account a certain escapement of material away from the immediate receiving environment by water transport or to allow for the concentration of contaminant X in the discharge and sediments to be similar.

Figure 3. Simple impact-distance relationship. This figure shows an exponential decrease in the effect on biota with distance from an impact source. The slope of such an impact curve is best achieved when samples are spaced (ΔX) so that the same change in impact (ΔY) is found between each pair.



The advantage of this type of calculation is that it should prevent samples from being collected in areas where it would be absolutely impossible to detect a change (or impact).

When to Sample

The timing and frequency of sampling are influenced by the goals of the monitoring program and the CMT. Some CMTs are insensitive to timing. For example, sediment sampling generally has a very small temporal variability. On the other hand, sampling in conjunction with bioaccumulation monitoring is critical because the reproductive cycles of marine organisms exert a major influence on the tissue concentrations of many contaminants. Consequently, if results are to be compared in a time series, samples must be taken at the appropriate time in order for compared results to be meaningful.

How to Sample

Many of the shortcomings of past effects-monitoring studies in the offshore Beaufort have been a result of a lack of awareness of the basic requirements of a sound environmental method and the limits imposed by arctic conditions. Effects studies involve a number of steps from planning to sample collection through analysis to interpretation. The end result is only as valid as the weakest link in the chain of operations. In the Beaufort region, extreme and prolonged cold for much of the year, continuous ice cover for at least eight months and the remoteness of the region from major population centres are all major obstacles to sample collection and analysis. As a result, sampling and field methods have often been the weakest link in Beaufort region environmental studies.

Proper application of sampling or measurement methodology is critical to the quality of monitoring data and ultimately to the conclusions drawn from the monitoring program. As pointed out in Taylor (1981) "Until a measurement operation has attained a state of statistical control, it cannot be regarded in any logical sense as measuring anything at all." We must be very careful, then, at each step of design and implementation to strive for recognising and controlling as many errors as possible. The hierarchy of methodology, proceeding from the general to the specific, may be considered as follows:

technique ------▸ method ------▸ procedure ------▸ protocol

A technique is a scientific principle that has been found to be useful for providing generic information.

A method is a distinct adaptation of a technique for a specific measurement purpose.

A procedure comprises the written directions necessary to use the method.

A protocol is the most specific name for a method. It is a set of definitive directions that must be followed, without exception, if the measurement results are to be accepted for a given purpose.

Monitoring programs should strive to use protocols because these will provide a benchmark or standard. If results are inadequate or ambiguous, for example, weaknesses in the protocol can be identified and improved. Improvements are always more effective when they are systematic because the improvements themselves fall into the realm of experimental science. Lack of protocols or failure to follow them compromises the acquisition of knowledge and slows progress.

As a general rule, no monitoring programs should be initiated until sampling and analytical methods, appropriate to the monitoring objective and having a proven history of satisfactory performance, have been chosen. The use of experimental techniques should be discouraged in routine monitoring programs, although they may be a valid or necessary component of a supplemental pilot program (to evaluate and intercalibrate new methods) or research program (to develop new methods) conducted within or outside the monitoring program.

One objective of the monitoring program should be to produce data which achieves a quality rating of 3 or 4 as defined by Thomas *et al.* (1983). Data achieving a rating of 3 are internally consistent. Patterns or trends within a data set are probably real, but comparison with other data sets may be a problem. Data achieving a rating of 4 are internally consistent and are sufficiently standardized or tied to a reference that comparison with other data at this rating should be possible. Although a 4 rating is obviously preferred, some types of measurements may be unable to achieve it because no effective means of standardisation or calibration may exist at the present time or because the measurements are subjective in nature. While 4 ratings are often possible for physical and chemical measurements, they are probably unachievable for measurements such as animal behaviourial responses.

Practical assistance in planning for, and obtaining, coherent data sets in studies involving the arctic marine benthic environment is provided in Arctic Laboratories Limited (1985). This report comprises two volumes. Volume 1 gives an overview of methods available for chemical and biological sampling of the arctic marine benthic environment. Volume 2 provides guidance in planning and choosing appropriate methods for a range of situations, conditions and objectives. The format of the report is designed to outline in a step-by-step manner, the factors to be considered and offers suggestions as to possible methods to achieve various results. Users are referred to other manuals and articles where detailed protocols are given.

Natural Variation

As emphasized throughout the preceding sections, addressing the question of natural variability in the CMTs will probably be the most important aspect of the design of the monitoring program. It is prudent, therefore, to subject the completed monitoring design to an adequacy test in the area of analysis of variance. A convenient way to do the evaluation is to check the monitoring program against the ten principles of impact monitoring design described by Green (1979). These principles deal primarily with the issue of variance in ecological systems.

1. Be able to state concisely to someone else what question you are asking. Your results will be as coherent and as comprehensible as your initial conception of the problem.

2. Take triplicate samples within each combination of time, location and any other controlled variable. Differences **among** can only be demonstrated by comparison to differences **within**.

3. Take an equal number of randomly allocated replicate samples for each combination of controlled variables. Putting samples in "representative" or "typical" places is not random sampling.

4. To test whether a condition has an effect, collect samples both where the condition is present and where the condition is absent but all else is the same. An effect can only be demonstrated by comparison with a control.

5. Carry out some preliminary sampling to provide a basis for evaluation of sampling design and statistical analysis options. Those who skip this step because they do not have enough time usually end up losing time.

6. Verify that your sampling device or method is sampling the population that you think you are sampling and with equal and adequate efficiency over the entire range of sampling conditions to be encountered. Variation in efficiency of sampling from area to area biases among-area comparisons.

7. If the area to be sampled has a large-scale environmental pattern, break the area up into relatively homogeneous subareas and allocate samples to each in proportion to the size of the subarea. If it is an estimate of abundance over the total area that is desired, make the allocation proportional to the number of organisms in the subarea.

8. Verify that your sample unit size is appropriate to the size, densities, and spatial distributions of the organisms that you are sampling. Then estimate the number of replicate samples required to obtain the precision you want.

9. Test your data to determine whether the error variation is homogeneous, normally distributed and independent of the mean. If this is not the case for most of the field data, then a) appropriately transform the data, or b) use a distribution free (non-parametric) procedure, or c) use an appropriate sequential sampling design, or d) test against simulated null hypothesis data.

10. Having chosen the best statistical method to test your hypothesis, stick with the results. An unexpected or undesired result is not a valid reason for rejecting the method and hunting for a "better" one.

Stage 5 Decision Point

At the completion of Stage 5, there are three possible outcomes:

1. abandon the null hypothesis as stated because it is impractical or impossible to test (e.g., unlikely to be able to resolve effect from natural variability). A solution may be to rephrase the null hypothesis into a form which can be tested and is worth testing, and reiterate Stage 5. If there is a lack of fundamental understanding about processes involving the CMT, then R + D may be recommended. It must be emphasised, however, that R + D has no role in monitoring programs. Recommendations for R + D must be done on a referral basis only.

2. if available information is insufficient or inappropriate, then pilot studies or preliminary studies may be recommended in order to acquire the

appropriate information to allow completion of the monitoring program design. Pilot studies are not part of a monitoring program *per se*, but are an integral part of program design. Pilot studies are often very important to test methods; this is particularly important in the Beaufort Sea because available literature may not be directly applicable.

3. if a monitoring program design has reached a satisfactory defensible conclusion, proceed to Stage 6.

STAGE 6. EXECUTE MONITORING PROGRAM

Stage 6 involves carrying out the monitoring program. No matter how good the design, problems will occur. Consequently, it is important to evaluate results and the performance of the program as quickly as possible so that implications for subsequent steps can be evaluated and corrective action taken if considered prudent. It must be remembered that each time a monitoring program is implemented, new information is produced which can be used to improve the next monitoring program and improve the evolving conceptual model of the effects of interactions between offshore hydrocarbon development and the Beaufort Sea receiving environment.

REPORTING

Reporting is an on-going process. It is an opportunity not just to report on monitoring results, but also to solicit comments and criticisms in order to improve the monitoring program and the results obtained. If possible, successive data collection phases of a monitoring program should not occur until data from a preceding phase has been analysed, reported and preferably evaluated. This evaluation is extremely important to all monitoring programs, but becomes that much more important in remote areas such as the Beaufort Sea where the cost of acquiring data is high. If changes to the sampling approach need to be made, then sooner is clearly better than later.

Reports must provide detailed protocols followed in each element of the monitoring program. Simple statements that "standard methods were followed" are not good enough. Systematic improvements can only be made when it is explicit as to what had been done. Quality control / quality assurance must accompany each element of the monitoring program. The report must include an assessment of the quality of the data; i.e., whether it is internally consistent (3 rating) or comparable to other data sets (4 rating). If the data fail to achieve the desired quality, then reasons must be sought and plans must be made to improve the data quality in subsequent stages of the program.

Peer review of interim reports should be part of the reporting process. This will allow an evaluation of interim results and provide feedback to the study design parameters before the study is complete.

DATA MANAGEMENT

Data management is a necessary component of the monitoring program. However, it should be designed and implemented with the view that data management should serve the monitoring program and make it more efficient; the monitoring program is not undertaken to serve the needs of data managers. The data management system should (1) be consistent over time; (2) use simple and cost-effective storage and retrieval methods; and (3) include information that will permit future quality evaluation of each data type.

RECOMMENDATIONS

1. A committee should be identified whose responsibility would be to review and approve monitoring program designs and methods.

2. At the present time, data gaps related to the natural variability of whale, bird and seal populations of the Beaufort Sea prevent proper design of monitoring programs involving these animals. No further effects-monitoring studies should be contemplated until the data gaps are filled.

3. The previous report entitled "Sampling and Analysis in the Arctic Marine Benthic Environment" should be updated to reflect recent advances in protocols particularly for benthic invertebrates.

4. Biological methods should be intercalibrated and steps taken immediately to improve the intercomparability of biological data sets. As biological measurements are more diverse and inherently more variable than chemical measurements, it is unlikely that biological data will ever attain the accuracy of chemical data. Nonetheless, improvements must be made to the way biological measurements are made in order to identify and understand relationships between environmental factors and biological responses.

5. Representatives from government, industry and local native groups should reach an *a priori* agreement on maximum acceptable effects level of environmental disturbance. This will lead to more effective and meaningful testing of impact hypotheses.

LITERATURE CITED

Addy, J.M., J.P. Hartley and P.J.C. Tibbetts. 1984. Ecological effects of low toxicity oil-based mud drilling in the Beatrice oilfield. Mar. Poll. Bull. 15 (12), 429-36.

Appelbee, J.F. and J.M. Mair. 1981. Hutton-Murchison fields, environmental baseline study, Volume II - macrofaunal assessment August 1980 Survey. Report prepared by the Institute of Offshore Engineering, Heriot-Watt University Edinburgh, U.K.

Arctic Laboratories Limited. 1985. Sampling and Ananysis in the Arctic Marine Benthic Environment. Volume 1. Review of methods. Volume 2. Guide to Practice. Prepared for Environment Canada, Environmental Protection Service, Yellowknife, N.W.T.

Battelle/Woods Hole Oceanographic Institution. 1983. Georges Bank benthic infaunal monitoring program. Final Report Year 1. U.S. Dept. of the Interior, Minerals Management Service. New York, U.S.A. 153 pp.

Birch, J.R., D.D. Lemon, D.B. Fissel and H. Melling. 1987. Arctic Data Compilation and Appraisal. Beaufort Sea and Almundsen Gulf: Physical Oceanography - Temperature, Salinity, Currents, Water Levels and Waves. Revised and updated to include 1914 through 1986. Can. Data Rep. Hydrography Ocean Sci. No. 5:459 pp.

Cornford, A.B., D.D. Lemon, D.B. Fissel, H. Melling, B.D. Smiley, R.H. Herlinveaux and R.W. Macdonald. 1982. Arctic Data Compilation and Appraisal. Volume 1. Beaufort Sea: Physical Oceanography -- Temperature, Salinity, Currents and Water Levels. Can. Data Rep. Hydrography Ocean Sci. No. 5:279 pp.

Davies, J.M., D.R. Bedborough, R.A.A. Blackman, J.M. Addy, J.F. Appelbee, W.C. Grogan, J.G. Parker and A. Whitehead. 1989. The environmental effect of oil-based mud drilling in the North Sea. pp. 59-89. (In) F.R. Engelhardt, J.P. Ray and A.H. Gillam (eds.), Drilling Wastes. Proceedings of the 1988 International Conference on Drilling

Wastes. Elsevier Applied Science. London and New York. 867 pp.

Davies, J.M., R.A.A. Blackman, J.R. Blanchard, J.E. Ferbrache, D.C. Moore, H.J. Somerville, A. Whitehead and T. Wilkinson. 1984. Environmental effects of the use of oil-based drilling muds in the North Sea. Mar. Pollut. Bull. 15 (10), 363-70.

Dome Petroleum Limited, Esso Resources Canada Limited and Gulf Canada Resources Inc. 1982. Environmental Impact Statement for hydrocarbon development in the Beaufort Sea-Mackenzie Delta Region. Volume 4. Biological and Physical Effects.

Duval, W.S., P. Vonk, M. Burnside and E. Einstein. (in prep). A procedure for determination of significance during environmental impact assessments. Seakem Group Limited. Vancouver.

Green, R.H. 1979. Sampling design and statistical methods for environmental biologists. John Wiley and Sons, Inc. New York. 257 pp.

Hargrave, B.T. and H. Thiel. 1983. Viewpoint: Assessment of pollution-induced changes in benthic community structure. Mar. Poll. Bull. 14(1), 41-46.

Harper, P.E. Jr., D.L. Potts, R.R. Salazer, R.J. Case, R.L. Jascheck and C.M. Walker. 1981. Distribution and abundance of macrobenthic and meiobenthic organisms. pp. 133-177 (In) B.S. Middleditch (Ed.), Environmental effects of offshore oil production. The Buccaneer Gas and Oil Field Study. Plenum Press, New York. 446 pp.

Harwood, L.A., L.A. Turney, L. de March, B.D. Smiley and P. Norton. 1986. Arctic Data Compilation and Appraisal. Volume 8. Beaufort Sea: Biological Oceanography -- Seals, 1826-1985. Can. Data Rep. Hydrography Ocean Sci. No. 5: Vol. 8 Part 1, 352 pp; Part 2, 301 pp.

Hoff, J.T. and D.J. Thomas. 1986. A compilation and statistical analysis of high quality

Beaufort Sea sediment data with recommendations for future data collections. Prepared by Arctic Laboratories Limited for Environment Canada (EPS). Yellowknife, N.W.T.

IMCO/FAO/UNESCO/WMO/WHO/IAEA/UN/UNEP. Joint Group of Experts on the Scientific Aspects of Marine Pollution. 1980. Monitoring biological variables related to marine pollution. Reports and Studies. No. 12, 21 pp.

INAC. 1986a. Summary report: Drilling fluid use and waste discharged in arctic marine waters of 60° for 1983. Arctic Waters Section. Water Resources Division. Indian and Northern Affairs Canada. Yellowknife. 38 pp.

INAC. 1986b. Summary Report: Drilling fluid use and waste discharge in arctic marine waters north of 60° for 1984. Arctic Waters Section. Water Resources Division. Indian and Northern Affairs Canada. Yellowknife. 47 pp.

INAC. 1985. Drilling mud usage and discharge in arctic marine waters north of 60° for the period 1978-1982. Arctic Waters Section. Water Resources Division. Indian and Northern Affairs Canada. Yellowknife. 6 pp. + Appendices.

Indian and Northern Affairs Canada. 1984-1988. Beaufort Environment Monitoring Project (BEMP). Environmental Studies Volumes 34, 39, 40, 52 and 60. Indian and Northern Afairs Canada. Ottawa.

Kingston, P.S. 1987. Field effects of platform discharges on benthic macrofauna. Phil. Trans. R. Soc. London. B316. pp. 545-65.

Mair, J.McD., I. Mathieson and J.F. Appelbee. 1987. Offshore macrobenthic recovery in the Murchison field following the termination of drill cuttings discharge. Mar. Pollut. Bull. 18 (12), 628-34.

Neff, J.M. 1987. Biological effects of drilling fluids, drill cuttings and produced waters.

pp. 469-538. (In) D.D. Boesch and N.N. Rabalais (eds.). Long-term environmental effects of offshore oil and gas development. Elsevier Applied Science. London and New York. 709 pp.

Norton, P., B.D. Smiley and L. De March. 1987. Arctic Data Compilation and Appraisal. Volume 10. Beaufort Sea: Biological Oceanography -- Whales, 1848-1983. Can. Data Rep. Hydrography Ocean Sci. No. 5:407 pp.

NRC. 1983. Drilling Discharges in the Marine Environment. National Academy Press. Washington, D.C., U.S.A. 180 pp.

Phillips, J.D.H. 1980. Quantitative Aquatic Biological Indicators. Applied Science Publishers. London.

Ratynski, R.A., L. de March and B.D. Smiley. 1988. Arctic Data Compilation and Appraisal. Volume 12. Beaufort Sea: Biological Oceanography -- Fish. 1896 to 1985. Can. Data Rep. Hydrography Ocean Sci. No. 5: Part 1, 301 pp., Part 2, 240 pp.

Reiersen, L-.O., J.S. Gray, K.H. Palmork and R. Lange. 1989. Monitoring in the vicinity of oil and gas platforms; Results from the Norwegian Sector of the North Sea and recommended methods for forthcoming surveillance. pp. 91-117 (In) F.R. Engelhardt, J.P. Ray and A.H. Gillam (eds.), Drilling Wastes. Proceedings of the 1988 International Conference on Drilling Wastes. Elsevier Applied Science. London and New York. 867 pp.

Taylor, J.K. 1981. Quality assurance of chemical measurements. Anal. Chem. 53 (14), 1588A.

Tetra Tech, Inc. 1987. Bioaccumulation monitoring guidance: Strategies for sample replication and compositing. Prepared for United Stages Environmental Protection

Agency, Office of marine and estuarine protection. Washington, D.C., U.S.A.

Thomas, D.J., G.D. Greene, W.S. Duval, K.C. Milne, and M.S. Hutcheson. 1983. Offshore oil and gas production waste characteristics, treatment methods, biological effects and their applications to Canadian regions. A report prepared by Arctic Laboratories Limited, ESL Enviromental Sciences Limited, SKM Consulting Ltd., and Atlantic Oceanics Company Ltd. for Environment Canada, Ottawa. 447 pp.

Thomas, D.J., R.W. Macdonald and A.B. Cornford. 1982. Arctic Data Compilation and Appraisal. Volume 2. Beaufort Sea: Chemical Oceanography. Can. Data Rep. Hydrography Ocean Sci. No. 5:243 pp.

Wainwright, P.F., B.D. Smiley and A. Blyth. 1987. Arctic Data Compilation and Appraisal. Volume 11. Beaufort Sea: Biological Oceanography -- Marine Zoobenthos, 1914 to 1986. Can. Data Rep. Hydrography Ocean Sci. No. 5:367 pp.

Wainwright, P.F. and L.B. Humphrey. 1988. Analysis of sediment data from the Beaufort Sea Shorebase Monitoring Program, 1982 to 1984. Environmental Studies Research Funds Report No. 090. Ottawa. 147 pp.

Woods, S. and B.D. Smiley. 1987. Arctic Data Compilation and Appraisal. Volume 9. Beaufort Sea: Biological Oceanography -- Bacteria, Plankton and Epontic Community, 1914 through 1985. Can. Data Rep. Hydrography Ocean Sci. No. 5:412 pp.

Yunker, M.B. 1985. Compilation of high quality benthic sediment and biota data sets for the Beaufort Sea. An unpublished report prepared by Dobrocky Seatech Ltd. for Environment Canada (EPS). Yellowknife, N.W.T.

APPENDIX A.

APPENDIX A.

Participants in the Workshop to Develop a Framework for Designing Effects Monitoring Studies in the Beaufort Sea.

Participant	Affiliation
William Cross	LGL Limited
Rolf Davis	LGL Limited
Wayne Duval	Seakem Group Ltd.
John Harper	Harper Environmental
Blair Humphrey	Seakem Group Ltd.
Peter McCart	Aquatic Environments Ltd.
Russel Shearer	Indian and Northern Affairs Canada
David Thomas	Seakem Group Ltd.