Our Reference No. (1-842)

Your Reference No. (06SB.KE603-5-0971)

October 15, 1986

Final Report
A Critical Literature Review of Representative Sampling and Subsampling of Heterogeneous Substances

by

Dr. Mark Yunker
Dobrocky Seatech Ltd.
9865 West Saanich Road
P.O. Box 6500
Sidney, B.C. V8L 4M7

for

Ocean Dumping Control Act
Environmental Protection Programs Directorate
Environment Canada
Ottawa, Ontario K1A 1C8

Scientific Authority: Rob Macdonald

DOBROCKY SEATECH

ACKNOWLEDGMENTS

First I would like to thank Dr. Byron Kratochvil of the University of Alberta in Edmonton for his assistance with this review of representative sampling and subsampling of heterogeneous substances. Dr. Kratochvil assisted with the update of the literature review and contributed Section 4.0 to this report.

I would also like to thank Dr. Rob Macdonald of the Institute of Ocean Sciences in Sidney, B.C. (the Scientific Authority) for his helpful comments and suggestions throughout the course of the review.

This project was funded under the Ocean Dumping Control Act Program of the Environmental Protection Programs Directorate of Environment Canada.

TABLE OF CONTENTS

LIST OF TABLES

1.0 INTRODUCTION

This critical literature review of representative sediment sampling and subsampling came about as the result of an unsolicited proposal submitted to the Ocean Dumping Control Act (ODCA) technical committee. The final terms of reference of the work were drafted to include a state-of-the-art review of the literature on the representative sampling and subsampling of heterogeneous substances. The results of this literature review were then summarized and conclusions reached on the application of the information obtained to sediment sampling.

Initially, relevant work was accessed by a key-word literature review and by assembling relevant papers collected by myself and by Drs. Kratochvil and Macdonald. When synthesizing the results of the literature review attention was paid to the following key aspects:

- the spatial distribution of the analyte in the matrix;
- random sampling error due to the heterogenity of particles; and,
- the influence of point sources and the stratification of pollutants on sampling.

This report is divided into a general literature review on the topic of sampling for chemical analysis, a section on the required number of samples for Ocean Dumping Control Act purposes and a section on the effects of sediment particle size. Finally the results of the critical literature review are summarized and recommendations are made as to what information is still required for administration of the Ocean Dumping Control Act.

2.0 GENERAL LITERATURE REVIEW

The topic of sampling for chemical analysis has recently been reviewed by Kratochvil et al. (1984). They performed a key word literature review on all aspects of sampling and it is unlikely that any references on the sampling of marine sediments were missed. Dr. Kratochvil updated the 1984 literature review as part of this work.

The Kratochvil et al. (1984) review was examined in detail and copies were obtained of all papers that appeared to have even peripheral relevance to the topic at hand. Very few of these papers had direct relevance to sediment sampling; many of them discussed procedures or concepts that were close to what was required, but in the end skirted the issue and could not be used. In this review, the intention is to discuss the key papers in a fair amount of detail, so that a clear statement can be made of what additional information is required to solve the existing problems in sediment sampling. For a complete list of papers on chemical sampling, readers are referred to Kratochvil et al. (1984).

2.1 SPATIAL DISTRIBUTION OF THE ANALYTE

Mackay and Paterson (1984) have addressed the issue of characterizing the heterogeneous spatial distribution of a chemical within a compartment or fraction of the environment using a probability density function. The advantage of this is the possibility of being able to make useful statements about exceedances -- that is, the fractions of the environment that have "excessive concentrations". The difficulties encountered, of course, are in converting the measurement data obtained into a distribution of concentrations, which when integrated, corresponds to the same total amount of chemical and in choosing the correct distribution function. Of the numerous functions available, they considered the normal, lognormal and Weibull functions and suggested that the Weibull function was particularly suitable as a distribution function because of its versatility. Indeed, it can closely reproduce both normal and lognormal distributions.

Like the lognormal distribution, the Weibull cannot give negative values, However, unlike the lognormal it can be integrated analytically to give a cumulative function which can be used to determine immediately the fraction of the total volume which experiences concentrations of less than or greater than a given value. In the example given (Mirex in Lake Ontario; Mackay and Paterson, 1984) the lognormal and Weibull distributions gave similar answers (the normal distribution was inappropriate since the standard deviation exceeded the mean). It appears that either distribution may be used depending on the exact goals of the data analysis. It should be noted that earlier workers (Quesenberry et al., 1976) have reported that the normal model fit aflatoxin data better than the lognormal model, and that the fit of the normal model was quite satisfactory.

Thompson and Howarth (1980) examined in detail the widely held view that the frequency distribution of analytical error was lognormal at concentrations of analyte near the detection limit. They argue that (a) there is no theoretical reason why such distributions should be lognormal and there is abundant evidence that they are not; (b) quasi-lognormal distributions can be produced as artifacts by data recording practices; and (c) inordinately large numbers of analytical results would be needed to distinguish a lognormal distribution from a normal distribution.

They do state, however, that the analysis of variance on log-transformed geochemical data is not likely to be rendered less valid by virtue of the more or less normal frequency distribution of analytical error. Only at near-detection limit concentrations are the differences between the normal and lognormal distributions marked and in any event, the effects of the violation of a different assumption (that log-transformed error is homoscedastic, i.e., independent of concentration) are likely to be greater.

Eberhardt et al. (1976) have commented on the frequency distribution of environmental data. They suggest that the common practice of assuming additive and normally distributed error seems to be neither empirically nor

logically acceptable. They reference data which show skewed frequency distributions and a relatively constant coefficient of variation and note that the gamma and lognormal frequency distributions are usually used for these data sets.

Eberhardt et al. (1976) then tested the consequences of assuming a lognormal distribution when the underlying distribution was in fact gamma (and vice versa) by testing the hypothesis that the coefficient of variation was in fact constant. F-tests using the two distributions performed in about the same manner, although tests on the lognormal distribution had a small but consistent advantage over tests on the gamma distribution. Similar but less conclusive results using the Scheffe's test for equality of variances suggested that it may not be particularly important whether the gamma or lognormal distribution actually holds -- log transformation will apparently serve many practical purposes in either case. It is noted that in excess of 200 observations are necessary to distinguish between the two distributions.

2.2 SAMPLING DESIGN

Garrett and Goss (1979) provided an improved method of evaluation of sampling and analytical variation in regional geochemical surveys. The method was based on an inverted nested sampling design followed by analysis of variance (ANOVA). The ANOVA partitioned the variability of the data into regional, sampling cell, lake and analytical components for center-lake bottom sediment surveys. Each region of interest was divided into 5 mi^2 cells. Once every 16 cells, a second lake within a cell was sampled; two samples were drawn from this lake, one of which was split to yield an analytical duplicate. The twentieth sample in each data block was reserved for one of several control reference samples used to monitor long-term analytical drift. Data were log transformed before analysis. This design had the desirable feature that the degrees of freedom for the analysis of variance were more evenly spread over all sampling levels than with a balanced design, and hence, more information on the regional variability was gained (at the expense of the lowest design stage) without increasing the overall size of the survey.

In addition to providing information on the variation at each of the four levels (region, cell, lake and analytical), it was also possible to determine if any of the components were not significantly different from zero and the partitioned variances could be used to establish confidence and prediction limits on the means. It should be noted that this scheme works best when analytical and within lake variations are not important contributors to the overall variation. Additionally, for the unbalanced design used, only approximate F tests are available to test hypotheses concerning the variance at each level. A more general approach to the sampling and interpretation of geochemical data has been provided by Miesch (1976), but most of the paper is peripheral to this discussion.

Eberhardt (1976) has discussed the key aspects of optimal sampling for parameter estimates. Probably the best known "descriptive" survey technique is that of stratified random sampling. Although stratification in both space and time is feasible, the major use in sampling for contaminants seems to be one of reducing variability due to spatial distribution. He comments that a judgemental choice of "representative" or "typical" sites can almost always be expected to lead to erroneous results. Also, since concentration weighted stratified random sampling tends to concentrate sampling in the "high level" strata, a grid scheme will not be as efficient in estimating total quantities of a contaminant.

Eberhardt (1976) noted that it was becoming fairly well recognized by 1976 that the frequency distributions of most environmental contaminants seldom resembled the convenient bell-shaped normal distribution but were instead "skewed" in shape, having a peak (modal) concentration at some relatively low level and a long tail trailing off to the high end of the scale. By 1976 two formal distributions had been proposed and used in radionuclide studies: the lognormal and the gamma distributions; they both have common properties (general shape and constant coefficients of variation) and yet differ considerably in mathematical properties and in models as to the mode of origin. The importance of frequency distributions depends very much on the objectives of the particular bit of statistical analysis at hand. In the case of sampling for estimating a <u>total quantity</u> of a contaminant

dispersed over some land area, one may largely ignore the shape of frequency distributions. If, however, statistical tests are in order, then it may become essential to consider the form of the frequency distribution (Eberhardt, 1976).

In calculating the required number of samples necessary for a sampling program, one can assume that the coefficient of variation is approximately constant, except in the presence of a substantial gradient in concentration.

Eberhardt (1978), has commented that a major share of the problems in studying trace substances in an ecological context appear to lie in the designing and analyzing of field studies. An essential preliminary for design is a clear statement of survey objectives. Of the four types of sampling discussed, descriptive sampling and sampling for spatial pattern are relevant to this discussion. Descriptive sampling covers what might be called the classical approach; the study objective is to get efficient estimates of a total quantity or, equivalently, an average or mean level. Stratified random sampling is usually used to obtain this total or mean (see discussion which follows). However, "to estimate efficiently a total or a mean" may not adequately describe the objectives of a sample survey in cases where the spatial pattern is regarded as important. We may instead want to learn a good deal about the location of the substance of interest. Eberhardt comments further that, inasmuch as concentrations of many trace substances exhibit skewed frequency distributions, for which the lognormal distribution appears to be a useful model, a suitable basis for contouring is systematic (grid) sampling and logarithmic transformation of the data (to approximate a constant error variance over the region of interest, without which the contouring process would seem to be inordinately complex). If such a setup is used, the contrast between sampling for a total (stratified random) and sampling for pattern (grid) can become very striking. Inasmuch as a concentration weighted stratified random sampling scheme tends to concentrate sampling in the "high level" strata, the grid scheme will not be very efficient in estimating total quantities of a contaminant (Eberhardt, 1976).

When the main objective is estimating a total quantity (of for example a radionuclide), the logical thing to do is to take single cores to a depth beyond which penetration is unlikely. The main problem is that, since such cores usually cannot be assayed in toto, subsampling must be used. Eberhardt (1978) comments that he believes that two subsamples should be taken from each core after mixing as a check on the effectiveness of the mixing process. If profile data are taken, but an estimate of the total for each location (equivalent to a single deep core) is also wanted, one can weigh the value of each incremental unit by a corresponding depth interval (i.e., a numerical integration). The main choice to be made is to decide which side of a compromise scheme gets the most emphasis -- whether to describe the profile with the usual incremental sampling or to estimate the total with composite cores.

Eberhardt et al. (1976) and Gilbert and Eberhardt (1976) have noted that "double sampling", in which a relatively inexpensive type of measurement is used in conjunction with some expensive chemical analysis, is an important tool in field sampling. In a sense, the costly but accurate method serves to "calibrate" a large sample in which the less expensive method is used. A total of N random locations can be selected for taking "fallible" measurements and a subset (n) of these locations would also be analyzed by the "true" method. The splitting of samples for the different types of analysis could occur at the field sample, field subsample or analytical digest stages.

Bernstein and Zalinski (1983) reviewed methods that provide a simple model for detecting a significant change in an environment and provided a power test procedure for determining the optimum sampling design for the detection of environmental impact. They emphasized that too often, conclusions about the presence or absence of an environmental impact (i.e., a significant biological or chemical change) are made based only on whether or not a statistically significant change has been detected. This approach is inappropriate because the size of detectable change depends greatly on the intensity of sampling (and for chemical species on the precision and

accuracy of the method). Thus, poorly funded and/or designed studies will detect no "significant" change, and studies with intensive sampling will find even small changes to be "significant".

At the outset, an estimate should be made of what constitutes a biologically (or chemically) significant impact, and then a sampling program with a specified probability (power) of detecting this change at a given level of statistical significance should be designed. Only when the probability and costs of making both type I and type II errors are made explicit is it possible to make an informed decision concerning the worth of proposed impact studies.

3.0 THE REQUIRED NUMBER OF SAMPLES

3.1 STATEMENT OF THE HYPOTHESIS TO BE TESTED

In Canada we have a defined limit L ug/g (0.6 ug/g for Cd and 0.75 ug/g for Hg) for a contaminant in a sediment above which we will require some evidence that the material is harmless or "rapidly rendered harmless" before permitting ocean dumping (Macdonald and Cretney, 1985). Of the two possible null hypotheses, the most suitable for Ocean Dumping Control Act purposes is the one where mean sediment concentration $\bar{x}$ in the area being evaluated is assumed to be greater than L and proof is required that the sediment is not contaminated and may be ocean dumped. The null hypothesis is then:

H_o : $\bar{x} \geq L$ (i.e., the sediment is contaminated)

H_A : $\bar{x} < L$ (i.e., the sediment is not contaminated)

and ocean dumping is prohibited if we accept H_o.

If the null hypothesis is stated in this way, a significance level alpha(1) can be chosen which will give a known probability of falsely rejecting the null hypothesis (a type I error). In this way we can specify exactly what risk we wish to take by choosing alpha(1) and prohibiting ocean dumping if we accept H_o. Standard practice is to choose alpha(1) = 0.05, although there are no hard and fast rules and the choice of alpha must depend on the perceived consequences of error. For example a smaller value for alpha(1) might be used for very large dredging projects. If H_o is accepted when it is indeed false (a type II error), then the environment is still protected. One may specify beta(1), the probability of committing a type II error, when calculating the required number of samples for a test (Zar, 1984, p. 110; Zar, 1981), but it normally can not be included when the hypothesis is being tested.

If the alternative hypothesis had been put forward (i.e., H_o: $\bar{x} \leq L$, the sediment is not contaminated), there would be an unknown risk of accepting contaminated material, since beta(1) can not be specified and the probability of accepting H_o when it was false would not be known. One would only be able to specify the probability (alpha(1)) of prohibiting ocean dumping unnecessarily.

If we assume (or know) that the log transformation is appropriate (Macdonald and Cretney, 1985) then the appropriate t statistic for a one-sided t test is:

$$t = \frac{(\ln L - \overline{\ln x})\sqrt{n}}{s} \qquad (3.1)$$

where s is the standard deviation associated with $\ln\bar{x}$ (the transformed mean) and n is the number of replicates. We reject H_o and permit ocean disposal when $t > t_{\text{alpha}(1),\, n-1}$.

3.2 EXAMINATION OF THE SOURCES OF VARIATION

The overall sample variance, s_o^2, is related to the variance of the sampling operation, s_s^2, and the variance of the remaining analytical operations, s_a^2, by the expression:

$$s_o^2 = s_s^2 + s_a^2 \qquad (3.2)$$

Whenever possible, measurements should be conducted in such a way that the components of variance arising from sample variability can be separately evaluated. When the measurement process has been demonstrated to be in a state of statistical control so that a reliable value can be assigned to s_a^2, then s_s^2 can be calculated from s_o^2, which is obtained from the actual analysis of samples from the area being studied or from a similar area.

The sampling variance, s_s^2, can also be broken down into a number of component variances. It can be regarded as being the sum of the between station variance, s_{st}^2, the between grabs at a station variance,

s_b^2, and the within grab at a station variance, s_w^2, as shown by the following:

$$s_s^2 = s_{st}^2 + s_b^2 + s_w^2 \text{ and} \tag{3.3}$$

$$s_o^2 = s_{st}^2 + s_b^2 + s_w^2 + s_a^2. \tag{3.4}$$

In the Ocean Dumping Control Act case, the survey objective is descriptive sampling; that is the objective is to get efficient estimates of the total quantity or, equivalently, an average or mean level (along with an estimate of the volume to be dredged) for the parameter of interest (Eberhardt, 1978). Since a mean will be calculated based on all samples collected, the variances between grabs at a station and within grabs at a station will usually not be examined and equation (3.2) is sufficient to characterize the overall variance. However, since the average or mean level will then be compared to the defined Ocean Dumping Limit L, it should be noted that it may become essential to consider the form of the frequency distribution of the contaminant being studied (Eberhardt, 1976).

Youden (1967, 1975) considered sampling to be a possible major source of error in the analytical process. He emphasized the importance of being able to place confidence limits on a result and that obtaining only a single result from a composite sample was of little value unless the variability of the parent population and analytical methodologies are known from extensive prior experience. As a guideline, he proposed that when the analytical error is one-third or less of the sampling error, further reduction of the analytical error is not important. If we assume that the analytical error is one-third of the sampling error, i.e.:

$$s_a = \frac{1}{3}s_s$$

then equation (3.2) can be restated in two ways:

$$s_o^2 = 10s_a^2 \tag{3.5a}$$

$$\text{or} \quad s_o^2 = \frac{10}{9}s_s^2 \tag{3.5b}$$

3.3 CALCULATION OF THE OPTIMUM NUMBER OF REPLICATES FOR ANALYSIS

The standard error (of the mean) is usually stated as (Zar, 1984, p. 87):

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

The standard error becomes a better estimate of the population standard error as n increases and $s_{\bar{x}}$ will directly influence the power of the t test used to test the hypothesis (see equation (3.1)).

For each site studied, a decision must be made about how to apportion relative effort between the sampling and analytical variances. While the power of the test increases with the number of replicate samples, there will be a level where the tests have sufficient power to test the hypotheses and the costs will be minimized. The most cost effective replicate structure will then be that which provides the maximum reduction in overall standard error for the minimum cost (Bernstein and Zalinski, 1983).

From equation (3.2), the estimated variance (of the overall mean) to be minimized is:

$$s_{\bar{\bar{x}}}^2 = \frac{s_s^2}{n_s} + \frac{s_a^2}{n_s n_r} \qquad (3.6)$$

where $s_{\bar{\bar{x}}}^2$ refers to the overall variance of the mean and n_s and n_r are the number of sediment samples and the number of analytical replicates respectively. For this discussion it is assumed that the number of analyses will be increased by a given percent (usually 10-20%) to include quality assurance samples (reference materials and reagent blanks) and that the cost of analysis of these control samples will be incorporated into the cost of the analyses. It is also assumed that a strictly hierarchical sampling scheme would be employed, with the result that the number of analytical replicates (n_r) is constant for the entire sampling and analysis program.

The total cost (C_T) of the sampling and analytical program can be expressed by:

$$C_T = n_s C_s + n_s n_r C_r$$

or alternatively

$$n_s = \frac{C_T}{C_s + n_r C_r} \tag{3.7a}$$

and

$$n_r = \frac{C_T - n_s C_s}{n_s C_r} \tag{3.7b}$$

where C_s and C_r are the costs of collecting a sample and analyzing an analytical replicate respectively.

Equation (3.6) can be restated as:

$$n_s s_{\bar{\bar{x}}}^2 = s_s^2 + \frac{s_a^2}{n_r}$$

Substituting for n_s (equation 3.7a) and rearranging terms we get:

$$n_r C_r s_s^2 + \frac{C_s s_a^2}{n_r} = C_T s_{\bar{\bar{x}}}^2 - C_s s_s^2 - C_r s_a^2$$

If we differentiate with respect to n_r, we find that the optimum number of analytical replicates is:

$$n_{r\,\text{optimum}} = \sqrt{\frac{C_s s_a^2}{C_r s_s^2}} \tag{3.8}$$

In actual practice, equation (3.8) may have to be used iteratively. While C_T will usually be fixed, C_s will depend on the number of samples that can be collected in the time available and C_r can vary with the number of samples analyzed.

In like manner, equation (3.6) can again be restated as:

$$n_r s_{\bar{\bar{x}}}^2 - \frac{n_r s_s^2}{n_s} = \frac{s_a^2}{n_s}$$

and on substituting for n_r (equation 3.7b) and rearranging terms we get:

$$n_s C_s s_{\bar{\bar{x}}}^2 + \frac{s_s^2 C_T}{n_s} = s_{\bar{x}}^2 C_T + C_s s_s^2 - C_r s_a^2$$

However on differentiating with respect to n_s, a very complicated equation was obtained for the number of samples n_s, that still contained C_T and $s_{\bar{\bar{x}}}^2$.

The variables required to calculate n_r optimum (equation 3.8) can all be obtained by independent means and hence a definitive value for n_r can be calculated. However, both C_T and $s_{\bar{\bar{x}}}^2$ are dependent on n_s. Even if C_T (the total program cost) is set at some maximum level, $s_{\bar{\bar{x}}}^2$ is still dependent on n_s and the result is that the equation for n_s optimum is a diverging function, and no matter what initial value is assumed for n_s (and hence $s_{\bar{\bar{x}}}^2$), n_s optimum will always be larger. Accordingly, this approach was not explored further.

3.4. CALCULATION OF THE NUMBER OF SAMPLES REQUIRED TO TEST THE HYPOTHESIS

In this case, the goal is to obtain enough representative samples of marine sediment from the area being examined, so that a valid test can be made of the null hypothesis (that the sediment is contaminated). The cost factor is brought into the discussion in Section 3.5. If the samples collected are not valid either due to contamination or due to the sample not being an unbiased representative of the population of interest, then the objectives of the study can not be met even if considerable care is taken to ensure outstanding analytical quality. Furthermore, if not enough samples are taken to give sufficient power to the statistical tests that will be applied to the analytical results, then the study objectives can not be met. Chemical sampling is discussed in detail by Kratochvil et al. (1984) and Kratochvil and Taylor (1981).

A relationship that may be used to calculate the required number of samples for a test against a given level for a given standard deviation (or variance) and a given acceptable error is:

$$n = \frac{t^2 s^2}{\bar{x}^2 R^2} \tag{3.9}$$

where n is the number of samples, t is the value from the standard t distribution based on the level of significance desired and the desired probability that a significant difference will be found (if it exists and is as small as the difference to be detected), s is the standard deviation, $\bar{x}$ is the mean and R is the coefficient of variation (relative standard deviation) acceptable in the average results (i.e., the minimum detectable difference (Zar, 1984, p. 110; Keith et al., 1983; Anonymous, 1984)). Note that this formula assumes that the underlying population is normally distributed and that the mean is larger than the variance (Kratochvil and Taylor, 1981).

For this discussion, a significance level of alpha(1) = 0.05 was chosen, which gives a 5% probability of falsely rejecting a null hypothesis that the mean is less than an defined limit (a Type I error). A probability of 90% of detecting a true difference between the mean and the defined limit as small as the difference to be detected (beta(1) = 0.01; i.e., the probability of commiting a Type II error and accepting the null hypothesis when it was false) was arbitrarily chosen. To use the formula t is initially set at 2.927 (t for alpha(1) = 0.05 plus t for beta(1) = 0.10, both at n = infinity; these are 1.645 and 1.282 respectively) and a preliminary value of n is calculated. The t values for this n are then substituted in the equation and the system is iterated to constant n. If the series does not converge, the higher predicted n is taken. If the number of samples predicted is less than two, values are rounded up to two since two is the minimum sample size allowable when using a t-test. This test is the minimum statistical test that may be applied to see if a significant change has occurred (Natrella, 1966).

3.5 EXAMPLES OF THE CALCULATION OF THE REQUIRED NUMBER OF SAMPLES

Cd and Hg data from the 1982 and 1983 samplings for the Canadian Beaufort Sea Shorebase Monitoring Program (Yunker, 1986; Thomas et al., 1983; Arctic Laboratories Ltd., 1984) were used to provide examples using equations (3.8) and (3.9). In 1982, sediment samples were collected in triplicate and analyzed from 67 subtidal stations (for a total of 201 samples) along the Tuktoyaktuk Peninsula and in Tuktoyaktuk Harbour and in Hutchison and McKinley Bays. A stratified random sampling design was used to choose the sampling locations in each embayment (Tuktoyaktuk Harbour and the two bays); an effort was made to select an equal number of sites in shallow and deep areas (relative to the 6 m isobath). Sampling locations outside of the embayments were chosen on a haphazard basis. In 1983, 74 samples were collected from Tuktoyaktuk Harbour, Hutchison Bay and McKinley Bay and from a few stations in the shallow Beaufort Sea. In both years of the program, a total (HF) digestion and graphite furnace atomic absorption was used for Cd and Hg was determined using a nitric, sulfuric, permanganate, persulfate digestion followed by cold vapour analysis. Cd levels varied from 0.01 - 0.93 ug/g and Hg from 0.001 - 0.14 ug/g. The data reported showed good precision, and were internally consistant and sufficiently standardized and tied to standard reference materials that they could be regarded as being of high quality.

An estimate of s_a^2, the analytical variance, was obtained by a pooling process using the results of the triplicate analyses of sediment homogenate that were carried out in both years of the program. The variances for each triplicate analysis were pooled using the formula given by Ku (1969). The rule is to combine the computed variances weighted by their respective degrees of freedom and then to divide by the total degrees of freedom for all samples. The pooled analytical variances are given below in ug/g dry weight.

Parameter	Pooled Analytical Variance	Mean	n
Cd	0.0005	0.29	20
Hg	0.00005	0.046	27

Because of the importance of Tuktoyaktuk Harbour, both as the major area of shorebase development and as a primary focus of the monitoring program, the Tuktoyaktuk Harbour stations were used to provide an estimate of the overall sample variance, s_o^2. The data from the 1982 subtidal sampling of Tuktoyaktuk Harbour were used and the mean and overall variance obtained for each parameter are given in Table 3.1. Since a large enough number of stations (12) and grab samples (36) were used to generate this Tuktoyaktuk data set, the variances obtained can be regarded as good estimates of the overall sample variance in Tuktoyaktuk Harbour. Assuming a normal distribution, these variances can then be used with confidence as a basis for calculating the number of samples required to produce a given minimum detectable difference (or coefficient of variation).

A comparison of the overall variance for each parameter in Table 3.1 with the pooled analytical variances given above shows that the analytical variances are small in comparison to the overall variances. This means that the overall variances will also provide a good estimate for the sampling variances, s_s^2, and can be used for the calculation of the minimum number of samples required to produce the minimum detectable differences (coefficients of variation) of (for example) less than 20%, 50%, 100% and 200% using the overall variances obtained from the Tuktoyaktuk Harbour samples.

In order to provide a "worst case" estimate of the required number of samples, the highest variance for each parameter that appeared in either the 1982 or 1983 data sets was used to calculate the minimum number of samples required to ensure a mimimum detectable difference of 20%, 50%, 100% and 200%. The results are also shown in Table 3.1. Even though the minimum number of samples required to produce minimum detectable difference of 20% is substantially higher for the "worst case" than for the Tuktoyaktuk Harbour samples, a detectable difference of 200% requires only one or two more samples for most parameters for the "worst case".

In Table 3.2, the analytical and sampling (both for Tuktoyaktuk Harbour and the "worst case") variances are given for Cd and Hg along with estimated "ball park" figures for the cost of an analysis (including Quality Assurance

Table 3.1a Estimates of the Minimum Number of Samples Required to Ensure a Minimum Detectable Difference of a Given Percent. Subtidal 1982 Tuktoyaktuk Harbour Data were used.

Parameter	Overall Mean ug/g	Overall Variance ug/g	n	Required Number of Samples to Produce a Coefficient of Variation of: 20%	50%	100%	200%
Cd	0.37	0.02	36	33	7	4	2
Hg	0.085	0.001	36	31	7	4	2

b Worst Case Estimates of the Minimum Number of Samples Required to Ensure a Minimum Detectable Difference of a Given Percent. Data for all of the 1982 and 1983 Subtidal Stations were Used.

Cd	0.235	0.043	275	167	29	9	4
Hg	0.061	0.0023	275	133	23	7	3

Table 3.2 Calculation of the Optimum Number of Analytical Replicates

Parameter	Symbol	Cd	Hg
Analytical Variance	s_a^2	0.0005	0.00005
Sampling Variance (Tuktoyaktuk)	s_s^2	0.02	0.001
Sampling Variance ("worst case")	s_s^2	0.043	0.0023
Cost of a Replicate Analysis (including Quality Assurance)	C_r	$45	$30
Cost of Collecting a Sample*	C_s	$45	$45
Optimum Number of Replicates (one minimum)	n_r optimum	1	1

* At 20 samples/day

samples) and for the cost of collecting a sample (assuming 20 grab samples/day and the use of a small vessel). It is also assumed that only short distances will be traveled in the collection of these samples. Application of equation (3.8) allows the calculation of the optimum number of replicates for analysis. Using the data given in Table 3.2, n_r optimum was calculated as being much less than one in all cases and hence a replicate size of one was specified.

As was mentioned in Section 3.3, it was not possible to derive an unambiguous formula to calculate the optimum number of samples. This means that the required number of samples must be calculated using a formula such as equation (3.9) and a number of possible solutions to the problem will result (see for example Table 3.1). The magnitude of difference from the Ocean Dumping Control Act Limit that can be detected for the parameters being studied (Cd and Hg) is a management decision and must be based on the risk and consequences of an incorrect decision and on the cost of collecting and analyzing the samples needed to make the decision.

4.0 THE EFFECT OF SEDIMENT PARTICLE SIZE

4.1 DETERMINATION OF SAMPLING CONSTANTS

The problem of sampling ocean or harbour sediments for contamination by trace levels of elements such as cadmium is complicated by several factors. These include:

1. A naturally occurring background level of the element, which may or may not be uniformly distributed in the sediments, must be taken into account. The element may be present in higher concentrations in some discrete grains or mineral forms in the sediment than in others, and the grains may not be well mixed in the area under study;

2. Superimposed upon this background may be highly variable concentrations of the element caused by anthropogenic activity such as dumping of wastes from industrial or municipal sources directly into the site as solids or in solution, or their introduction by streams or rivers. The element under study will likely be present at higher concentrations in some particles of the foreign material than in others. Also, it may be absorbed from solution on the surfaces of particles. In this case it would be present to a larger extent on the smaller particles, where the surface area is high relative to volume, and on particles where specific interactions are favored, such as the ion exchange of metals onto clays;

3. Information on the distribution of contaminants such as the trace element cadmium amoung various types and sizes of particles typically found in ocean sediments is rare; and,

4. A variety of mixing and segregating processes may affect sediment distributions relative to point source contamination. Currents, tides, winds, and wave action may all affect distribution patterns.

Lacking much of this basic information, one of two general approaches may be made to setting up a program for assessing the level of an element such as cadmium in the sediments of a particular region of the ocean floor. In the first approach as much information as possible is collected on the distribution of the element in similar systems. Assumptions are made as to how closely the available data fit the area of interest, and where data are lacking, further assumptions are made based on experience, judgement and intuition. Using this data set the number, size and locations of the sampling increments required to give a specified level of sampling precision are calculated from statistical principles. For example, if strata can be identified, a stratified sampling plan can be used; if the distribution of the sought-for component is Gaussian, the number of increments can be estimated from the relation:

$$n = \frac{t^2 s^2}{R^2 \bar{x}^2} \times 10^4 \qquad (4.1)$$

where t is the student's t-table value for the level of confidence desired, the variance s^2 and average percentage composition $\bar{x}$ are estimated, and R is the percent relative standard deviation (coefficient of variation) acceptable in the average. Initially t is set at 1.960 for 95% confidence limits and a preliminary value calculated for n. The t value for this n is then obtained from statistical tables and the system iterated to constant n (Kratochvil and Taylor, 1981).

In the second approach no prior assumptions about the population distribution of the substance of interest are made. A set of sample increments is collected from the site of interest under the assumption that heterogeneity is appreciable, and are analyzed for the substance. On the basis of the analytical results a sampling plan is devised that will give the desired precision.

In practice a combination of the two approaches is best. As much information as possible is collected about the likely distribution of the

element to be determined. From this information, along with appropriate assumptions where information is not available, a preliminary sampling plan is drawn up. The more preliminary information available, the less extensive the preliminary sampling operation need be.

The three parameters that must be specified in the sampling of a population such as ocean floor sediments for an element such as cadmium are the size of the increments to be collected (and shape if variation of the element with depth in the sediment must be evaluated), the number of increments to be collected, and the locations from which they are to be obtained. Initially it will be assumed that the population of interest is a top layer, say 10 cm, of sediment on the ocean floor that is uniform in cadmium content with respect to depth, and that the sampling device collects a layer of relatively uniform thickness over a specified area of ocean floor. We will also assume for the estimates here that a grab sampler of the VanVeen type is employed for sample collection, and that an increment of 1 to 1.5 kg wet weight of sediment is taken from the centre of the grab. The increment will approximate after drying a cube 5 cm on a side, and will weigh approximately 500 g.

Under these conditions the standard deviation of the sampling operation s_s can be best obtained by use of the Visman relation:

$$s_s^2 = A/wn + B/n \qquad (4.2)$$

where w is the weight of an individual increment, n is the number of increments, and A and B are constants determined by the distribution of cadmium in the sediment (Visman, 1969, 1972; Visman et al., 1971). The size of constant A depends on the heterogeneity within local increments, and so is affected by the size of the increments taken as well as the number. The size of constant B depends on the variability of the concentration on a large scale as, for example, when the local concentration of cadmium is increased by the dumping of a contaminated batch of material onto the ocean floor.

There are two ways of estimating values for A and B. In the first, two sets of increments are collected, one set having an increment weight of w_{sm} and the second an increment weight of w_{lg}. The size of the increments is determined by local heterogeneity and so must be estimated in this preliminary step, but a value for w_{sm} of 0.5 to 1 g is suggested for sediments in which particle diameters are of the order of 1 mm or less. The value for w_{lg} should be at least ten times the value of w_{sm}. A set of 20 to 30 increments of each size should be collected. If information is not available concerning the location of areas of high cadmium contamination at (for example) a dump site, then the increments should be collected from stations located on a systematic grid pattern, with the initial site being determined by some random process. If, on the other hand, information on regions of possibly high cadmium concentration is available, a stratified sampling plan can be set up. This requires knowledge of the area of the high cadmium region, however, and also assumes that within-strata regions are uniform in cadmium concentration. Because neither areas nor uniformity are likely to be known with adequate precision and accuracy, the use of stratified sampling is not considered to be practical in this case.

Once the large and small increments have been analyzed for cadmium the standard deviation of the sampling operations is calculated for each by:

$$s_s^2 = s_o^2 - s_a^2 \tag{4.3}$$

where s_o is the standard deviation of the overall analytical process, and s_a is the standard deviation of the analytical measurement operations (Youden, 1967). Then values for A and B are calculated by:

$$A = \frac{w_{lg} \; w_{sm}}{w_{lg} - w_{sm}} (s_{sm}^2 - s_{lg}^2) \tag{4.4}$$

and

$$B = s_{lg}^2 \frac{A}{w_{lg}} \tag{4.5}$$

With knowledge of A and B it is possible to calculate values for w and n that will hold the sampling standard deviation to any desired level.

A second method for the estimation of A and B arises from the theory of cluster sampling. Here a series of pairs of increments are collected, each member of a pair being of the same weight w and collected from nearby sites. From analyses of the increments an interclass correlation coefficient r is calculated by

$$r = [(p-1)s_b^2 - ps_w^2]/[(p-1)s_b^2 + ps_w^2] \tag{4.6}$$

where s_b^2 and s_w^2 are the between-set and within-set variances obtained from an analysis of variance (ANOVA) of the data, and p is the number of pairs of data used in the calculation. Once the value of r is known, values for A and B are obtained by simultaneous solution of the equations:

$$B = s^2 \frac{A}{w} \tag{4.7}$$

and

$$r = \frac{mB}{A} \tag{4.8}$$

Here m is the average particle mass, and must be estimated for the population. From the data of Clague (1976), sediments collected near Comox B.C. ranged from 95% gravel to 42% clay and 54% silt in the inner trough. The median values for 108 samples in the inner trough were 0% gravel, 15.3% sand, 55.5% silt, and 25.6% clay. Using the Wentworth size class, where silt ranges from 4 to 62 um in diameter, a conservative value of 60 um will be used here for the average particle diameter. Assuming a density of 2.6 for the sediment, the average particle mass is then

$\frac{2.6}{6} \pi d^3$, or 0.29 ug.

If more than 20 to 30 increments are thought to be necessary to provide the desired sampling precision, but the cost of analysis is appreciably greater than the cost of collecting the initial increments, then consideration should be given to the collection of additional increments and partially compositing them to give 20 to 30 composite samples. Under these conditions a modified form of Visman's equation should be used (Wallace and Kratochvil, unpublished research). For analyses of composites consisting of y small increments of weight w_1, or x large increments of weight w_2, selected so that $yw_1 = w_{sm}$, and $xw_2 = w_{lg}$, A and B in Visman's equation are given by:

$$B = y\ s_{sm}^2 - \frac{A}{w_1} = x\ s_{lg}^2 - \frac{A}{w_2} \tag{4.9}$$

and

$$A = \frac{w_2 w_1}{w_2 - w_1} (y\ s_{sm}^2 - x\ s_{lg}^2) \tag{4.10}$$

Whether compositing should be done depends on the relative cost of the sampling and analysis operations at the time of the study.

Under the conditions described here sample locations are best sited on an equispaced grid over the harbor area to be studied. To maintain the element of randomness necessary to allow statistical treatment of the resulting data, the initial point on the grid should be selected by some random process. The other points can then be spaced relative to the starting point (Kratochvil and Taylor, 1981). A grid sampling plan has the advantage that all regions of the population are sampled, and it is simpler to do; error is introduced only when the concentration of sought-for substance varies in a periodic way in phase with the sampling spacing. This is unlikely for a harbor sediment.

Two factors affect the size of sample increments to be collected. One is the size of the tool available for the sampling operation. The second is the size of the increments needed to determine the Visman constants A and B. The following section discusses increment size from another point of view.

4.2 HETEROGENEITY AND PARTICLE SIZE

Sometimes the size of the increment collection device does not permit collection of large and small increments in the range needed to calculate A and B. In such cases other approaches may be used to estimate increment size. One way is to assess local variability in distribution of the component of interest in the particles comprising the population. This contribution may be significant even though the population of particles is well mixed, i.e., homogeneous at the grain level. The relevant mathematical relationships were worked out by Benedetti-Pichler (1956) in the 1950's. His conclusion was that the significant parameters were the variability in composition of the material among particles, and the number of particles collected for analysis. For simplicity he considered the bulk material to be a two- component mixture, with each component containing a different percentage of analyte. Thus, for a sediment containing trace cadmium, one type of particle might contain some level of cadmium P_1 in percentage, and the second type of particle essentially no Cd ($P_2 = 0$). The number of particles n required to hold the relative sampling standard deviation (the sampling uncertainty) R in percent to a preselected level may be calculated from the relation

$$n = \left[\frac{d_1 d_2}{\overline{d}^2}\right] \left[\frac{100(P_2 P_1)^2}{R\overline{P}}\right] (p)(1-p) \qquad (4.11)$$

where d_1 and d_2 are the densities of the two kinds of particles, $\overline{d}$ is the average density of the sample, $\overline{P}$ is the overall average composition in percent of the component of interest in the sample, and p and 1-p are the fractions of the two kinds of particles in the bulk material. If the particle size of a material is small, then n will be large for a given sample weight and the value of R will be small. If the particles are large, then either a larger sample must be used, or the material must be ground to increase n.

For example, suppose a sample of a sediment of the general composition of that near Comox, B.C. (see earlier) is to be collected. Values of n, along

with the assumptions used for the various factors in equation (4.11), are given in Table 4.1. From Table 4.1 it can be concluded that for a material of 60 um diameter (approximately 250 mesh), the test portion size required to hold the sampling error to 5% relative can vary widely, depending on the assumptions made as to the fraction of the total particles that contain high levels of contaminant, and the concentration of contaminant in the two types of particles modelled. If the contaminant is assumed to be cadmium at an overall average concentration ($\overline{P}$) of 1 ug/g, then the size of the portion required for analysis varies as shown in Table 4.2. Here it can be seen that the subsample weight must be 10 g or greater if the Cd contaminant is concentrated in a few particles, even though the general background level is a tenth of the total average Cd level.

The overall conclusion is that if overall cadmium (or other) contamination is present at concentrations of 1 ug/g (1 ppm) then very large subsamples are required if the cadmium is present in only a limited number of particles. The subsample size will be reduced substantially if the particles are smaller than 60 um, or if the contaminant is dispersed over a larger number of particles. However, the assumptions used here provide an initial guideline, and a warning that sample increment size should be carefully evaluated in preliminary sampling studies to ensure that it is large enough to avoid the introduction of error from statistical particulate patterns.

Table 4.1 Weights of test portions of a well mixed sample that are required to hold the coefficient of variation in sampling to 5% for mixtures of two types of 60 um diameter spherical particles of density 2.6.

Calculations are for several levels of % trace contaminant in the major kind of particle present. (i.e. p_1% Cd in the p fraction of the total number of particles required) and for several higher levels of the contaminant in the minor kind of particle present (i.e., P_2% in the 1-p fraction of particles).

P_1	P_2	p	1-p	$\bar{P}$	n	
% contamination in matrix fraction	% contamination in minor fraction	fraction of particles containing P_2% contaminant	fraction of particles containing P_1% contaminant (matrix fraction)	Av. % contaminant in material	No. of particles required	Wt. of test portion required, in g
0	1	0.01	0.99	0.010	4.0×10^4	0.012
1×10^{-6}(0.01 μg/g)	"	"	"	"	4.0×10^4	0.012
1×10^{-4}(1 μg/g)	"	"	"	"		
0	1	1×10^{-3}	0.999	1.0×10^{-3}	4.0×10^{-5}	0.12
1×10^{-6}	"	"	"	"	"	"
1×10^{-5}	"	"	"	"	"	"
1×10^{-4}	"	"	"	1.1×10^{-3}	3.3×10^{-5}	0.099
0	1	1×10^{-4}	1	1.0×10^{-4}	4.0×10^6	0.99
1×10^{-6}	"	"	"	"	"	"
1×10^{-5}	"	"	"	1.1×10^{-4}	3.3×10^6	0.99
1×10^{-4}	"	"	"	2.0×10^{-4}	1.0×10^6	0.30
0	1	1×10^{-5}	1	1×10^{-5}	4.0×10^7	12
1×10^{-6}	"	"	"	1.1×10^{-5}	3.3×10^7	9.9
1×10^{-5}	"	"	"	2.0×10^{-5}	1.0×10^7	3.0
1×10^{-4}	"	"	"	1.1×10^{-4}	3.3×10^5	0.099

Table 4.1 (continued)

P_1	P_2	p	1-p	$\bar{P}$	n	
% contamination in matrix fraction	% contamination in minor fraction	fraction of particles containing P_2% contaminant	fraction of particles containing P_1% contaminant (matrix fraction)	Av. % contaminant in material	No. of particles required	Wt. of test portion required, in g
0	1	1×10^{-6}	"	1.0×10^{-6}	4.0×10^{8}	120
1×10^{-6}	"	"	"	2.0×10^{-6}	1.0×10^{8}	30
1×10^{-5}	"	"	"	1.1×10^{-5}	3.3×10^{6}	0.99
1×10^{-4}	"	"	"	1.0×10^{-4}	4.0×10^{4}	0.012
0	0.01	0.01	0.99	1.0×10^{-4}	4.0×10^{8}	120
1×10^{-5}	"	"	"	1.1×10^{-4}	3.3×10^{8}	99
1×10^{-4}	"	"	"	2.0×10^{-4}	1.0×10^{8}	30
0	10	1×10^{-6}	1	1.0×10^{-5}	4.0×10^{6}	1.2
1×10^{-6}	"	"	"	1.1×10^{-5}	3.3×10^{6}	0.99
1×10^{-4}	"	"	"	1.1×10^{-4}	3.3×10^{4}	0.010

Table 4.2 Size of test portion required to hold s_s to 5% relative for a sediment containing 1 ppm (1 µg/g) Cd.

P_1		P_2		p^a	$1-p$	
Cd in matrix fraction		Cd in contaminant fraction		fraction of particles that are contaminated	fraction of particles comprising matrix	wt. of test portion required, g[b]
%	µg/g	%	µg/g			
0	0	0.1	10^3	1×10^{-5}	1	1.16×10^1
1×10^{-8}	0.01	"	"	1×10^{-5}	1	1.16×10^1
1×10^{7}	0.1	"	"	9×10^{-6}	1	1.04×10^1
0	0	1	10^4	1×10^{-6}	1	1.16×10^2
1×10^{-8}	0.01	1	10^4	"	"	"
1×10^{-7}	0.1	1	10^4	9×10^{-7}	"	1.04×10^2
0	0	10	10^5	1×10^{-7}	1	1.16×10^3
1×10^{-8}	0.01	10	"	"	1	1.16×10^3
1×10^{-7}	0.1	10	"	9×10^{-8}	1	1.04×10^3
0	0	100	10^6	1×10^{-8}	1	1.16×10^4
1×10^{-8}	0.01	"	"	"	1	1.16×10^4
1×10^{-7}	0.1	"	"	9×10^{-9}	1	1.04×10^4

[a]For this table $p = \frac{\bar{P} - P_1}{P_2 - P_1}$.

[b]$wt = (n \text{ particles})(0.29\ \mu g/\text{particle})$; $n = (p)(1-p)\left[\frac{100(P_2-P_1)}{5(1 \times 10^{-6})}\right]^2$.

5.0 SUMMARY OF RESULTS AND RECOMMENDATIONS

5.1 SUMMARY OF RESULTS

When the results of this literature review are examined, a number of glaring information gaps remain. Most obvious is the fact that there is very little relevant data of documented high quality on the spatial distribution of the analyte (Cd or Hg) in the sediment matrix. In many cases total metal concentrations are given along with the grain size distribution but there is almost no data on the variation of metal concentration within size class. As a result of this lack of information it is not possible to state whether (a) large enough samples or (b) sufficient samples are being analyzed to satisfy the goals of the Ocean Dumping Control Act (ODCA). Furthermore, the lack of information on the frequency distribution of Cd and Hg in sediments may imply that the statistical tests used to test conformance of a sediment to the Act (before ocean dumping) may not be valid.

The general literature review pointed out the fact that the importance of frequency distributions depends very much on the objectives of the particular bit of statistical analysis at hand. In the case of estimating a total quantity of a contaminant dispersed over the area being sampled, one may largely ignore the shape of frequency distributions. If, however, statistical tests are in order (as they are for ODCA administration), then it may become essential to consider the form of the frequency distribution (Eberhardt, 1976). The review also pointed out that a judgmental choice of "representative" or "typical" sites can almost always be expected to lead to erroneous results. The best known "descriptive" survey technique is that of stratified random sampling. A grid scheme is not as efficient in estimating total quantities of a contaminant, however, it is usually necessary for contouring.

For ODCA purposes, the most suitable null hypothesis to test is $H_o : \bar{x} \geq L$ (ie, the sediment is contaminated). When H_o is stated in this way it is possible to specify exactly what risk we wish to take in testing H_o since it will be possible to give a known probability of falsely rejecting the null hypothesis and concluding that the sediment is not contaminated. If the alternative hypothesis had been put forward (ie, $H_o : \bar{x} \leq L$, the sediment is not contaminated), one would only be able to specify the probability of prohibiting ocean dumping unnecessarily.

Based on the costs of collecting and analyzing a sample and on a knowledge of the sampling and analytical variances, a formula was put forward for calculating the optimum number of analytical replicates that should be analyzed. In most cases it is likely that the formula will predict one replicate. Essentially then, once an analytical procedure has been demonstrated to be in a state of statistical control, it is best to concentrate resources on determining the sampling variation (ie to analyze samples from different grab or core samples as opposed to analytical replicates). This assumes that an adequate number of quality assurance samples will also be analyzed concurrently. The Geological Survey of Canada (Garret and Goss, 1979) uses essentially the same approach with an inverted nested sampling design, where most of the samples analyzed are from different sampling locations.

For ODCA testing, enough representative samples of marine sediment need to be collected from the area being examined, so that a valid test can be made of the null hypothesis (that the sediment is contaminated). From a knowledge of the mean and the variance, it is possible to calculate the number of samples required to determine any minimum detectable difference from the OCDA limit. It is, however, not possible to optimize the number of samples to be collected for any given total cost of a program. The decision on the number of samples to be analyzed has to be based on the magnitude of difference from the OCDA limit desired (ie the minimum detectable difference), the risk and consequences of an incorrect decision and finally on the cost of the program. Once these decisions have been made, the formulae and examples shown allow this number of samples to be easily calculated.

The results of this literature review also point out an information gap in the area of sampling constants for the analytes of interest. A knowledge of Visman sampling constants is necessary to allow the calculation of the weight of individual samples and the number samples to hold the sampling standard deviation to a given level. Alternatively, the sampling constants allow an assessment of what the sampling standard deviation will be for a given weight of sample and a given number of samples (based on both the calculated number of samples given above and the cost).

Sampling theory demonstrates that for a material of 60 um diameter (approximately 250 mesh), the test portion size required to hold the sampling error to 5% relative can vary widely depending on the assumptions made as to the fraction of the total particles that contain high levels of contaminant, and the concentrations of contaminant in the two types of particles modelled. The overall conclusion is that if the metal contamination being studied (for example, Cd or Hg) is present at concentrations of about 1 ug/g (1 ppm), then very large subsamples are required if the metal is present in only a limited number of particles.

5.2 RECOMMENDATIONS

From the above discussion, it follows that more information is required on the distribution of the analytes of interest to the Ocean Dumping Control Act (Cd and Hg) in marine sediments from typical dredge sites. It is recommended that a number of sites be chosen on each coast and enough samples collected to determine the frequency distribution by sediment size class and the Visman sampling constants for Cd and Hg in the sediments at each site. It is expected that significant cost savings would be realized in this work by making the collection and analysis concurrent with the ODCA application process at as many of the sites as possible.

6.0 REFERENCES

Anonymous, 1984. Workshop report on the status and trends program: Recommendations for design and implementation of the chemical measurements segment. Produced by the Ocean Assessment Division, National Oceanic and Atmospheric Administration, Rockville Md., 39 p.

Arctic Laboratories Ltd., 1984. Beaufort Sea Coastal Sediment Reconnaissance Survey: A data report on 1983 geochemical sampling. An unpublished report prepared for the Environmental Protection Service, Yellowknife, Northwest Territories.

Benedetti-Pichler, A.A. In Physical Methods in Chemical Analysis; Berl. W.G. Ed.,; Academic Press, New York, 1956, Vol. III, p 183-217.

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

Clague, J.J., 1976. Sedimentology and Geochemistry of Marine Sediments near Comox, B.C. Geol. Surv. Can. Pap. 76-21, 26 p.

Eberhardt, L.L., 1978. Designing ecological studies of trace substances. In Environmental Chemistry and Cycling Processes. Proceedings of a symposium held at Augusta GA April 28 - May 1, 1976. Published by Technican Information Center, U.S. Dept. Energy, p. 8-33.

Eberhardt, L.L., 1976. Sampling for radionuclides and other trace substances. In Radioecology and Energy Resources (C.E. Cushing Jr., ed.) The Ecological Society of America, Spec. Publ. No. 1. Dowden, Hutchinson and Ross Inc., Stroudsberg Penn. p. 199-208.

Eberhardt, L.L., R.O. Gilbert, H.L. Hollister and J.M. Thomas, 1976. Sampling for contaminants in ecological systems. Environ. Sci. Technol. 10: 917-925.

Garrett, R.G. and T.I. Goss, 1979. The evaluation of sampling and analytical variation in regional geochemical surveys. Geochem. Explor., Proc. Int. Geochem. Explor. Symp., 7th: 371-383.

Gilbert, R.O. and L.L. Eberhardt, 1976. An evaluation of double sampling for estimating plutonium inventory in surface soil. In Radioecology and Energy Resources (C.E. Cushing Jr., ed.). The Ecological Society of America, Spec. Publ. No. 1. Dowden, Hutchinson and Ross Inc., Stroudsberg, Penn. p. 157-163.

Keith, L.H., W. Crummett, J. Deegan, R.A. Libby, J.K. Taylor and G. Wentler, 1983. Principles of environmental analysis. Anal. Chem. 55: 2210-2218.

Kratochvil, B. and J.K. Taylor, 1981. Sampling for chemical analysis. Anal. Chem. 53: 924A-938A.

Kratochvil, B., D. Wallace and J.K. Taylor, 1984. Sampling for chemical analysis. Anal. Chem. 56: 113R-129R.

Ku, H.H., 1969. Precision measurement and calibration, statistical concepts and procedures. U.S. National Bureau of Standards Special Publication 300, Volume 1, Washington, D.C. p. 316.

Macdonald, R. and W. Cretney, 1985. Regulated limits: The Ocean Dumping Control Act (Canada) as an example. Unpublished document from the Ocean Chemistry Division, Institute of Ocean Sciences, Sidney, B.C., 16 p.

Mackay, D. and S. Paterson, 1984. Spatial concentration distributions. Environ. Sci. Technol. 18: 207A-214A.

Miesch, A.T., 1976. Geochemical survey of Missouri -- methods of sampling, laboratory analysis and statistical reduction of data. U.S. Geol. Surv. Prof. Paper 954-A, 39 p.

Natrella, M.G., 1966. Experimental statistics. National Bureau of Standards Handbook 91, Washington, D.C.

Quesenberry, C.P., T.B. Whitaker and J.W. Dickens, 1976. On testing normality using several samples: An analysis of peanut aflatoxin data. Biometrics 32: 753-759.

Thomas, D.J., P.F. Wainwright, B.D. Arner, and W.H. Coedy, 1983. Beaufort Sea Coastal Sediment Reconnaissance Survey: A data report on 1982 geochemical and biological sampling. An unpublished report prepared for the Environmental Protection Service, Yellowknife, Northwest Territories.

Thompson, M. and R.J. Howard, 1980. The frequency distribution of analytical error. Analyst, 105: 1188-1195.

Visman, J., 1969. A general sampling theory. Mat. Res. Stds., 9: 9-13, 51-56, 62-64.

Visman, J., A.J. Duncan and M. Lerner, 1971. Further discussion: a general theory of sampling. Mat. Res. Stds., 11:32-37.

Visman, J., 1972. A general theory of sampling, Discussion 3. J. Mat. 7: 345-350.

Wallace, D., Kratochvil, B., 1986, unpublished research. Unpublished research from the Department of Chemistry, University of Alberta, Edmonton, Alberta.

Youden, W., 1967. J. Assoc. Off. Anal. Chem., 50: 1007.

Youden, W.J., 1975. Statistical techniques for collaborative tests. In statistical manual of the Association of Official Analytical Chemists. Published by the Association of Official Analytical Chemists, Arlington, Va.

Yunker, M. B., 1986. 1982-83 Beaufort Sea Shorebase Monitoring Program; Statistical Analysis and Recommendations for Future Programs. An unpublished report prepared for Indian and Northern Affairs Canada, Yellowknife Northwest Territories, 190 p.

Zar, J.H., 1981. Power of statistical testing: hypotheses about means. Amer. Labor, 103-107.

Zar, J.H., 1984. Biostatistical analysis. Prentice-Hall Inc., Eaglewood Cliffs, New Jersey.