PNL-4804
1-
NUREG/CR-4162
PNL-4804

Survey of Statistical and Sampling Needs for Environmental Monitoring of Commercial Low-Level Radioactive Waste Disposal Facilities

Prepared by L. L. Eberhardt, J. M. Thomas

Pacific Northwest Laboratory
Operated by
Battelle Memorial Institute

Prepared for
U.S. Nuclear Regulatory
Commission

REFERENCE COPY

NOTICE

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, or any of their employees, makes any warranty, expressed or implied, or assumes any legal liability of responsibility for any third party's use, or the results of such use, of any information, apparatus, product or process disclosed in this report, or represents that its use by such third party would not infringe privately owned rights.

NOTICE

Availability of Reference Materials Cited in NRC Publications

Most documents cited in NRC publications will be available from one of the following sources:

1. The NRC Public Document Room, 1717 H Street, N.W.
Washington, DC 20555
2. The Superintendent of Documents, U.S. Government Printing Office, Post Office Box 37082,
Washington, DC 20013-7082
3. The National Technical Information Service, Springfield, VA 22161

Although the listing that follows represents the majority of documents cited in NRC publications, it is not intended to be exhaustive.

Referenced documents available for inspection and copying for a fee from the NRC Public Document Room include NRC correspondence and internal NRC memoranda; NRC Office of Inspection and Enforcement bulletins, circulars, information notices, inspection and investigation notices; Licensee Event Reports; vendor reports and correspondence; Commission papers; and applicant and licensee documents and correspondence.

The following documents in the NUREG series are available for purchase from the GPO Sales Program: formal NRC staff and contractor reports, NRC-sponsored conference proceedings, and NRC booklets and brochures. Also available are Regulatory Guides, NRC regulations in the *Code of Federal Regulations*, and *Nuclear Regulatory Commission Issuances*.

Documents available from the National Technical Information Service include NUREG series reports and technical reports prepared by other federal agencies and reports prepared by the Atomic Energy Commission, forerunner agency to the Nuclear Regulatory Commission.

Documents available from public and special technical libraries include all open literature items, such as books, journal and periodical articles, and transactions. *Federal Register* notices, federal and state legislation, and congressional reports can usually be obtained from these libraries.

Documents such as theses, dissertations, foreign reports and translations, and non-NRC conference proceedings are available for purchase from the organization sponsoring the publication cited.

Single copies of NRC draft reports are available free, to the extent of supply, upon written request to the Division of Technical Information and Document Control, U.S. Nuclear Regulatory Commission, Washington, DC 20555.

Copies of industry codes and standards used in a substantive manner in the NRC regulatory process are maintained at the NRC Library, 7920 Norfolk Avenue, Bethesda, Maryland, and are available there for reference use by the public. Codes and standards are usually copyrighted and may be purchased from the originating organization or, if they are American National Standards, from the American National Standards Institute, 1430 Broadway, New York, NY 10018.

Survey of Statistical and Sampling Needs for Environmental Monitoring of Commercial Low-Level Radioactive Waste Disposal Facilities

Manuscript Completed: March 1985
Date Published: July 1986

Prepared by
L. L. Eberhardt, J. M. Thomas

Pacific Northwest Laboratory
Richland, WA 99352

Prepared for
Division of Radiation Programs and Earth Sciences
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555
NRC FIN B2461

Survey of Statistical and Sampling Needs for Environmental Monitoring of Commercial Low-Level Radioactive Waste Disposal Facilities

Manuscript Completed: March 1985
Date Published: July 1986

Prepared by
L. L. Biedrzycki, J. M. Thomas

Pacific Northwest Laboratory
Richland, WA 99352

Prepared for
Division of Radiation Programs and Earth Sciences
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC 20555
NRC File B261

ABSTRACT

This document presents the results of Task 1 of a project entitled "Application of Statistics in Siting and Managing Commercial Low-Level Radioactive Waste Disposal Sites." Task 1 was designed to develop guidance for determining the overall needs for sampling and statistical work in characterizing, surveying, monitoring, and closing commercial low-level waste sites. The overall project is designed to produce information for developing guidance on implementing 10 CFR Part 61.

Results are listed in the Executive Summary and areas requiring additional investigation are discussed in Chapter 5.0 (Summary and Conclusions). If cost-effectiveness and statistical reliability are considered of prime importance, then double sampling, compositing, and stratification (with optimal allocation) are identified as key issues for NRC's consideration. Alternatively, if the principle concern is avoiding questionable statistical practice, then the applicability of kriging (for assessing spatial pattern), methods for routine monitoring, and use of standard textbook formulae in reporting monitoring results should be reevaluated. Other important issues are identified in the report.

ABSTRACT

This document presents the results of Task 1 of a project entitled "Application of Statistics in Siting and Managing Commercial Low-Level Radioactive Waste Disposal Sites." Task 1 was designed to develop guidance for determining the overall needs for sampling and statistical work in characterizing, surveying, monitoring, and closing commercial low-level waste sites. The overall project is designed to produce information for developing guidance on implementing 10 CFR Part 61.

Results are listed in the Executive Summary and areas requiring additional investigation are discussed in Chapter 5.0 (Summary and Conclusions). If cost-effectiveness and statistical reliability are considered of prime importance, then double sampling, composition, and stratification (with optimum allocation) are identified as key issues for DOE's consideration. Alternatively, if the principle concern is avoiding questionable statistical practice, then the applicability of kriging (for assessing spatial pattern), methods for routine monitoring, and use of standard textbook formulas in reporting monitoring results should be reevaluated. Other important issues are identified in the report.

EXECUTIVE SUMMARY

The goal of this report is to provide a survey of statistical and sampling needs for environmental monitoring of low-level radioactive waste disposal facilities. Four classes of sampling (descriptive, analytical, pattern and sampling for modeling) have been investigated for application to LLW sites. Our objective was to provide enough details and examples to permit an intelligent choice of the methods that will be most appropriate for the main applications anticipated by NRC staff and user groups. In addition, this report can serve in part as a reference document for NRC staff on sampling for radioactivity and/or monitoring in areas other than low-level waste management, as well as providing information for developing guidance on implementing 10 CFR Part 61. The actual adaptation to specific field uses is the subject of further investigation.

A detailed summary, which includes a list of research recommendations, is in Chapter 5.0. Its purpose is to provide a broad view of research topics identified in the current project. In contrast, with a few exceptions, this Executive Summary points out results, areas where minimal research might lead to recommendations and identifies areas where research has already begun. Each of the following summary statements is accompanied by a reference to the appropriate report section.

Predictive Value of Computer Simulation Models:

We believe that complicated computer simulation models will not generally have good predictive value and that simpler linear and non-linear regression models will produce more realistic projections, since they are generally fit to actual field data. Supporting arguments are in Section 2.2.

Stochastic Versus Deterministic Models:

When coefficients of variation are near unity (as in the case of trans-uranics) models should be stochastic as opposed to deterministic. If not, results which are supposed to be "in the ballpark" may vary by orders of magnitude (Section 2.3).

Invalid Statistical Analyses:

Because the underlying statistical model which describes most LLW site radionuclide monitoring data is not normal and additive, many formulae taken from standard textbooks cannot be directly used. Statistical analyses which are wrong can result (Section 2.3).

Compositing - Complications:

Compositing samples (for estimation or detection) from LLW sites can be both cost effective and complicated. For instance, mean values for composites can be under or overestimated by 1.7 when lognormality of samples which make up the composite is not accounted for (Section 3.2).

Compositing - Allocation of Sampling Effort:

When radioanalyses are expensive, compositing of samples may be necessary. However, allocation of sampling effort is quite different when the question is determining an average value in soils at a LLW site as compared to attempting to detect the location of a spill (compare Figures 3.4 and 3.5).

Compositing - Estimating a Site Mean or Total:

When compositing to estimate a mean or total quantity at LLW sites, it appears that a routine procedure (after a thorough preliminary experiment) might be to run radioanalyses of two subsamples of a single composite, but to periodically run two composites with a single subsample of each (Section 3.3).

Compositing - Detecting Spills or Migration:

Based on our review of compositing for detection, we conclude that this technique may be very useful in detecting offsite migration or spills at future LLW sites. We are investigating a scheme for selecting a "best" sampling strategy for a particular location which could be run on the kind of micro computer likely to be available to NRC staff, state regulatory agencies, or site operators. In order to evaluate feasibility in a timely manner, we will not consider some important unresolved statistical issues in the pilot study (Section 3.4).

Optimal Sample Weight or Volume:

Additional research on optimal sample weight or volume for components of composites as well as subsamples is needed, since our review shows that larger sample masses or volumes substantially reduce the coefficient of variation. However, the gain can be offset by a larger mixing variance (Section 3.5).

Censored Distributions (Light Bulbs Versus Less Than Detectable Radioactivity):

The considerable literature available concerning right-censored distributions (i.e., failure times for light bulbs and animal survival times) should be explored for applicability to monitoring data from LLW sites, which are generally from left-censored distributions (e.g., less than detection values are often reported and can seriously compromise statistical evaluation (Table 3.2)). In addition, methods for statistically comparing data from two left-censored distributions should either be obtained from the literature or developed (Section 3.6).

Wasted Sampling Effort:

Allocation of field sampling effort (where and how many samples) is objective dependent. Poor allocation can result in wasted resources (Section 4.1).

Use of Survey Data to Answer Questions:

Survey data can be used in the analytic mode, e.g. has there been a significant change? Features peculiar to LLW-site monitoring data (inferences about cause can be complicated, errors are usually multiplicative and some samples will exhibit values below detection limits) should be anticipated (Section 4.1.2).

Estimating Model Parameters:

When sampling to estimate parameters in a mathematical model of a process (e.g., radionuclides in soil as a function of distance from source), evenly spaced sampling increments will generally be inefficient (Sections 4.1.3 and 4.4.1).

Spatial Pattern:

When sampling for spatial pattern of radionuclides, the choice of sample locations and frequencies will depend on the availability of a mathematical model of the process governing movement. When a model is available, the methods in Section 4.4 should be considered, otherwise the methods of geostatistics in Section 4.5 may apply (Section 4.1.4).

Systematic Versus Random Sampling:

Systematic sampling can be recommended over completely random or stratified sampling when small-scale patterns exist (i.e., there is a demonstrable spatial correlation among nearby sites; see Section 4.2.2 for details). However, spatial periodicity (e.g., roads, fence rows, etc.) can result in a biased result when systematic samples are taken (Section 4.2.1).

Routine Monitoring:

Routine sampling (e.g., routine air sampling at a fixed place and fixed time intervals) may be uninformative simply because of a poor choice of location or sampling frequency. We recommend that different times and places be considered unless detecting change is the main motive for the study (Section 4.2.6).

Cost Reduction:

Opportunities for reducing costs will sometimes depend on the fact that a LLW-site manager has special knowledge that some event of concern may occur or has occurred (e.g., spill or loss of radionuclide from confinement). Especially cost effective designs or techniques in this case include stratified sampling, ratio and regression estimation (double sampling) and cluster sampling and subsampling. Examples for LLW sites are in Sections 4.2.3 through 4.2.5. Since the matter of allocation and design of stratified sampling is extremely important, we are investigating this further. In addition, research has been initiated on the use of sample sizes smaller than 30 for double sampling.

Sample Size:

Knowledge of the coefficient of variation can be a useful guide in selecting the appropriate sample size for LLW site studies (Section 4.2.7).

Comparison of Means from Several Monitoring Locations:

Results from more than two subplots (stations) must often be evaluated for statistically significant differences (assuming replicate or more than one subsample per station). Our review shows that there are no appropriate multiple comparison methods for the unbalanced (different number of subplots/station) and heterogeneity of variance (variances among stations differ) case which is likely to be common at low level waste sites. For all other cases, it appears that a computer program which could produce graphical output to judge the statistical significance of station differences would be useful for future low-level waste site managers (Section 4.3.2).

Identification of Causes of Radionuclide Movement:

The "pseudodesign" approach may be applicable for low-level waste sites to identify the cause of radionuclide redistribution and/or migration (Section 4.3.4).

Uniform Versus Other Spacing of Sample Collections:

When sampling to estimate parameters in mathematical models of a process (e.g., as a function of time, distance, or depth), uniform spacing of sample collections (e.g., once a day, one meter intervals, or five centimeter depths) is usually a bad choice (Section 4.4.1).

Kriging:

We believe that the basic assumptions of kriging (a statistical technique for estimating spatial pattern) must be thoroughly investigated for applicability to low-level waste management data. Our arguments supporting this statement are in Section 4.5.6 while some examples of the uses of kriging are in Section 4.5.7.

ACKNOWLEDGMENTS

We acknowledge the contribution of John Skalski for much of the material in Section 3.4. Marji Cochran provided the computerized bibliography, Gail Poole prepared the final manuscript and Rene Hinds edited an earlier version.

ACKNOWLEDGMENTS

We acknowledge the contribution of John Skafel for much of the material in Section 3.4. Mark Cochran provided the computerized bibliography. Carl Boole prepared the final manuscript and Rene Hinds edited an earlier version.

CONTENTS

	<u>Page</u>
ABSTRACT.....	iii
EXECUTIVE SUMMARY.....	v
ACKNOWLEDGEMENTS.....	ix
 1.0 INTRODUCTION.....	 1-1
1.1 OBJECTIVES OF THE PROJECT.....	1-1
1.2 SCOPE OF NRC'S STATISTICAL REQUIREMENTS.....	1-1
1.3 APPROACH AND FUTURE WORK.....	1-7
 2.0 BACKGROUND.....	 2-1
2.1 INTRODUCTION.....	2-1
2.2 MODELS.....	2-5
2.2.1 Projections Versus Forecasting.....	2-6
2.2.2 Deterministic, Stochastic and Simulation Models..	2-7
2.2.3 Classes of Models.....	2-8
2.3 FREQUENCY DISTRIBUTIONS AND TRANSFORMATIONS.....	2-8
 3.0 SAMPLING UNITS.....	 3-1
3.1 INTRODUCTION.....	3-1
3.2 COMPOSITING.....	3-1
3.3 COMPOSITING FOR ESTIMATION.....	3-3
3.4 COMPOSITING FOR DETECTION.....	3-5
3.5 OPTIMAL VOLUME OR WEIGHT FOR SAMPLES.....	3-7
3.6 CENSORED DATA SETS.....	3-7
 4.0 SAMPLING METHODS.....	 4-1
4.1 INTRODUCTION.....	4-1
4.1.1 Descriptive (Survey) Sampling.....	4-2
4.1.2 Sampling for Comparisons (Analytical Sampling)...	4-2
4.1.3 Sampling for Modeling.....	4-4
4.1.4 Sampling for Pattern.....	4-4
4.2 DESCRIPTIVE (SURVEY) SAMPLING.....	4-4
4.2.1 Introduction.....	4-4
4.2.2 Systematic Sampling.....	4-5
4.2.3 Cost-Effective Survey Sampling Methods.....	4-7
4.2.4 Further Details on Stratified Sampling.....	4-9
4.2.5 Double-Sampling.....	4-11
4.2.6 Repeated Sampling in Time.....	4-12
4.2.7 Sample Size.....	4-12
4.2.8 References on Descriptive Sampling.....	4-13

	<u>Page</u>
4.3 ANALYTICAL SAMPLING.....	4-17
4.3.1 Introduction.....	4-17
4.3.2 Multiple Comparisons.....	4-18
4.3.3 Statistical Evaluation of Data Collected for Analytical Purposes.....	4-21
4.3.4 The Question of Cause.....	4-25
4.3.5 References on Analytical Sampling.....	4-26
4.4 SAMPLING FOR MODELING.....	4-28
4.4.1 Introduction.....	4-28
4.4.2 The Box and Lucas Scheme.....	4-29
4.4.3 Errors of Measurement.....	4-33
4.4.4 Additive and Multiplicative Parameters.....	4-33
4.4.5 Discriminating Between Models.....	4-34
4.4.6 Response Surface Models.....	4-34
4.4.7 References on Sampling for Modeling.....	4-35
4.5 SAMPLING FOR PATTERN.....	4-36
4.5.1 Introduction.....	4-36
4.5.2 Trend Surfaces.....	4-37
4.5.3 Moving Averages.....	4-39
4.5.4 Spatial Splines.....	4-40
4.5.5 Stochastic Process Models.....	4-40
4.5.6 Kriging.....	4-41
4.5.7 Some Applications of Kriging.....	4-44
4.5.8 References on Sampling for Spatial Patterns.....	4-45
5.0 SUMMARY AND CONCLUSIONS.....	5-1
6.0 REFERENCES CITED.....	6-1

FIGURES

	<u>Page</u>
2.1 Three classes of study designs.....	2-2
2.2 A comparison of sampling for worldwide fallout and undisturbed sampling at low-level waste sites.....	2-4
3.1 Compositing of samples can be used as a means of keeping the costs of analyses down.....	3-2
3.2 Hypothetical example of compositing for estimation.....	3-3
3.3 Hypothetical example of compositing for detection.....	3-6
4.1 Four classes of sampling at a hypothetical low-level waste shallow land burial facility.....	4-3
4.2 Subsampling plots at a hypothetical low-level waste shallow land burial site.....	4-19
4.3 Tritium concentration in water samples obtained at four Maxey Flats stations as a function of time.....	4-23
4.4 A hypothetical example of cesium-137 concentration as a function of soil depth which shows uniform versus geometric sampling for an exponential curve.....	4-30
4.5 A possible consequence of fitting trends with a polynomial model.....	4-38

FIGURES

4-38	polynomial model.....	4-38
4-30	geometric sampling for an exponential curve.....	4-30
4-24	A possible consequence of fitting trends with a function of soil depth which shows uniform versus	4-24
4-23	A hypothetical example of cesium-137 concentration as a	4-23
4-19	Maxey Flats stations as a function of time.....	4-19
4-13	tritium concentration in water sample obtained at four	4-13
4-12	shallow land burial site.....	4-12
4-7	Subsampling plots at a hypothetical low-level waste	4-7
4-3	shallow land burial facility.....	4-3
3-6	Four classes of sampling at a hypothetical low-level waste	3-6
3-3	hypothetical example of composting for detection.....	3-3
3-2	hypothetical example of composting for estimation.....	3-2
3-2	the costs of analysis alone.....	3-2
3-1	Composting of samples can be used as a means of keeping	3-1
2-4	undisturbed sampling at low-level waste sites.....	2-4
2-2	A comparison of sampling for worldwide fallout and	2-2
2-2	three classes of study designs.....	2-2

1.0 INTRODUCTION

1.1 OBJECTIVES OF THE PROJECT

The overall objective of the project under which this report was written is to do the research necessary to provide guidance for the use of statistics in planning, conducting surveys, monitoring and assessing data relative to the siting, operation and decommissioning of low-level waste disposal sites. In the Statement of Work, which initiated this project, the Nuclear Regulatory Commission staff asked for "an initial assessment of the substantial theoretical information and experience that may be adapted for NRC needs." The present report seeks to provide that assessment. Since it is an initial assessment, we have not attempted to reach any final conclusions, beyond re-emphasizing some of the features that we believe need to be adapted for NRC needs. We note that "NRC needs" are very diverse, ranging from the specifics of closing particular existing sites (such as Maxey Flats or Sheffield) to the generic issues of providing guidance for selecting, operating, and closing sites in the future. Statistical and sampling requirements will undoubtedly differ greatly in the two situations, but hopefully dealing with present problems will yield useful insights for planning for the future and in developing guidance for implementing 10 CFR Part 61.

The purpose of this report is to describe a general framework for sampling in low-level waste management, and the main elements of the following four different classes of sampling methods:

1. descriptive or survey sampling, in which the objective is to estimate a total quantity or an average;
2. analytical sampling, directed primarily towards making comparisons between subsets of the population being sampled;
3. sampling for modelling, in which efforts are focused on estimating parameters for some mathematical model; and
4. sampling for pattern, where the objective is the efficient gathering of data for representing a spatial pattern.

The project goal is to provide enough details and examples of applications of the four approaches to permit an informed choice of the methods that will be most appropriate for the main applications anticipated by NRC staff and user groups. The actual adaptation to specific field uses requires further investigation.

1.2 SCOPE OF NRC'S STATISTICAL REQUIREMENTS

In order to emphasize those areas where investigation would be most useful to NRC, we reviewed various NRC documents as well as the rule-making provided in 10 CFR 61 and several reports and drafts dealing with low-level waste site characterization, selection, and monitoring. Excerpts from 10 CFR 61 and the U.S. Nuclear Regulatory Commission Branch

Technical Position on Site Suitability, Selection and Characterization (NUREG-0902) follow, to show examples where we believe various sampling and statistical approaches relate directly to NRC rules and guidelines. In all cases the citations are direct quotations from the respective documents.

The format for listing specific quotations includes the sections of 10 CFR 61 or page numbers from the Branch Technical Position where we believe statistical treatment of data is implied (design, hypothesis testing, analysis, data collection, etc.), a sometimes long quotation (so that it is not out of context), followed by an assessment (in the Branch Technical Position) of the statistical implications (in most cases titles of sections of our report are used). We seldom specifically mention the material discussed in chapters 2.0 or 3.0 of the report because they must be considered in almost all cases. The listing of monitoring citations from 10 CFR 61 does not give sections of our document which address monitoring, since the specific section or sections would depend on the question being asked (e.g., how much? where? is there a difference over the site? does the concentration observed exceed a prescribed limit?).

Finally, we note these are not exhaustive listings. We only wish to give the reader an overall impression of the numbers and kinds of statistical issues involved in low-level waste disposal.

SECTIONS OF 10 CFR PART 61 WHERE STATISTICAL DESIGN, ANALYSIS OR
MODELLING IS IMPLIED OR REQUIRED (CODE OF FEDERAL REGULATIONS, 1982)

Section	Quotation
61.7 (c) (4)	"After a finding of satisfactory disposal site closure, the Commission will transfer the license, to the state or Federal government that owns the disposal site. If the Department of Energy is the Federal agency administering the land on behalf of the Federal government, the license will be terminated because the Commission lacks regulatory authority over the Department for this activity. Under the conditions of the transferred license, the owner will carry out a program of monitoring to assure continued satisfactory disposal site performance, physical surveillance to restrict access to the site and carry out minor custodial activities. ..."
61.12 (k)	"A description of the radiation safety program for control and monitoring of radioactive effluents to ensure compliance with the performance objective in 61.41 of this part and occupational radiation exposure to ensure compliance with the requirements of Part 20 of this chapter and to control contamination of personnel, vehicles, equipment, buildings and the disposal site. Both routine operations and accidents must be addressed. The program description must include procedures, instrumentation, facilities and equipment."
61.12 (l)	"A description of the environmental monitoring program to provide data to evaluate potential health and environmental impacts and the plan for taking corrective measures if migration of radionuclides is indicated."
61.13 (a)	"Pathways analyzed in demonstrating protection of the general population from releases of of radioactivity must include air, soil, ground-water, surface water, plant uptake and exhumation by burrowing animals. The analyses must clearly identify and differentiate between the roles performed by the natural disposal site characteristics and design features in isolating and segregating the wastes. The analyses must clearly demonstrate that there is reasonable assurance that the exposure to humans from the release of radioactivity will not exceed the limits set forth in 61.41."
61.28 (a) (2)	(Application for closure must include) "The results of tests, experiments, or other analyses relating to back-fill of excavated areas, closure and sealing, waste migration and interaction with emplacement media, or any

Section	Quotation
	other tests, experiments, or analysis pertinent to the long-term containment of emplaced waste with the disposal site."
61.29	"Following completion of closure authorized in 61.28, the licensee shall observe, monitor, and carry out necessary maintenance and repairs at the disposal site until the license is transferred by the Commission in accordance with 61.30. Responsibility for the disposal site must be maintained by the licensee for 5 years. A shorter or longer time period for post-closure observation and maintenance may be established and approved as part of the site closure plan, based on site-specific conditions."
61.41	"Concentrations of radioactive material which may be released to the general environment in ground surface water, soil, plants, or animals must not result in an annual dose exceeding an equivalent of 25 millirems to the whole body, 75 millirems to the thyroid, and 25 millirems to any other organ of any member of the public. Reasonable effort should be made to maintain releases of radioactivity in effluents to the general environment as low as is reasonably achievable."
61.53 (a)	"At the time a license application is submitted, the applicant shall have conducted a preoperational monitoring program to provide basic environmental data on the disposal site characteristics. The applicant shall obtain information about the ecology, meteorology, climate, hydrology, geology, geochemistry and seismology of the disposal site. For those characteristics that are subject to seasonal variation, data must cover at least a twelve month period."
61.53 (c)	"During the land disposal facility site construction and operation, the licensee shall maintain a monitoring program. Measurements and observations must be made and recorded to provide data to evaluate the potential health and environmental impacts during both the construction and the operation of the facility and to enable the evaluation of long-term effects and the need for mitigative measures. The monitoring system must be capable of providing early warning of releases of radionuclides from the disposal site before they leave the site boundary."
61.53 (d)	"After the disposal site is closed, the licensee responsible for the post-operation surveillance of the disposal

STATISTICAL ISSUES IMPLIED IN U. S. NUCLEAR REGULATORY COMMISSION BRANCH
TECHNICAL POSITION ON SITE SUITABILITY, SELECTION AND CHARACTERIZATION
(NUREG-0902; SIEFKIN ET AL., 1982)

Section	Quotation
	site shall maintain a monitoring system based on the operating history and the closure and stabilization of the disposal site. The monitoring system must be capable of providing early warning of releases of radionuclides from the disposal site before they leave the site boundary."
61.55 (a) (8)	"Determinations of concentrations in wastes. The concentration of a radionuclide may be determined by indirect methods such as use of scaling factors which relate the inferred concentration of one radionuclide to another that is measured, or radionuclide material accountability, if there is reasonable assurance that the indirect methods can be correlated with actual measurements. The concentration of a radionuclide may be averaged over the volume of the waste, or weight of the waste if the units are expressed as nanocuries per gram."
61.80 (i) (2)	"The reports shall include (i) specification of the quantity of each of the principal radionuclides released to unrestricted areas in liquid and in airborne effluents during the preceding year, (ii) the results of the environmental monitoring program, (iii) a summary of licensee disposal unit survey and maintenance activities, (iv) a summary, by waste class, of activities and quantities of radionuclides disposed of, (v) any instances in which observed site characteristics were significantly different from those described in the application for a license; and (vi) any other information the Commission may require. If the quantities of radioactive materials released during the reporting period, monitoring results, or maintenance performed are significantly different from those expected in the materials previously reviewed as part of the licensing action, the report must cover this specifically."

Page	Quotation	Statistical Implications
1	"The NRC is also responsible to assure that these commercial nuclear facilities comply with the National Environmental Policy Act (NEPA) of 1969."	Analytic Sampling
3	"The disposal site shall be capable of being characterized, modeled, analyzed, and monitored."	Descriptive and Analytic Sampling. Sampling for Pattern and Modeling
5	"Nearby facilities or activities" ...cannot ...significantly mask the environmental monitoring program."	Quantitative methods to separate site sources from other sources of radiation. Descriptive Sampling.
5	"Since site characterization investigations can sample only a small fraction of the surface area or subsurface volume of the disposal site, site characteristics must be such that these limited investigations can adequately define the site characteristics spatially across the disposal site. Site conditions should be such that well-documented analytical solutions or computer codes are available and applicable for modeling site performance."	Sampling for Pattern. (NRC perhaps will accept results from simulation models.)
5	"Finally, since monitoring programs can sample only a small fraction of the surface area or subsurface volume of the disposal site, site characteristics must be such that the small number of monitoring points can adequately describe the extent to which radionuclides have migrated from the waste disposal units."	Sampling for Modeling. Analytical Sampling. Compositing.
8	"In addition, this requirement should provide sufficient space within the buffer zone to implement remedial measures, if needed, to control releases of radionuclides before discharge to the ground surface or migration from the disposal site."	Compositing. Analytical Sampling. Sampling for Modeling.

Page	Quotation	Statistical Implications
9	"In cases of co-located facilities, the monitoring programs for the facilities should be able to identify the source and differentiate between the releases of the separate facilities. In addition, the monitoring programs should provide adequate data to identify mitigative measures if needed, for the separate facilities."	Descriptive Sampling. Sampling for Pattern. Analysis of Ratios (Not addressed in this report.)
11	"The evaluation of available reconnaissance-level data during the site selection process should serve as a basis for design of the site characterization program." (... "a program of investigations and tests, both in the field and laboratory" ...).	Design.* (All report sections)
12	"Specific objectives of site characterization are to develop the technical information needed for... "5) establishment of data collection points and a baseline of data for some portions of the site monitoring program, and" ...	Design.
13	(NRC and applicant discussions on defining)... "size of study area, the depth of borings, the hydrogeologic units to be investigated, and points of interest for determination potential releases of radionuclides."	Design.

*It appears to us that the site characterization phase (pages 11-24, NUREG/0902) not only is to provide information about current status of ground and surface water, soil characterizations, and air quality (and other measurements), but also that the information (e.g., the sampling frequency and location of sampling stations) will carry over to the monitoring program during operation and closure. The implication in some of the following quotations is that a change in parameter values based on the site characterization period compared to the operating period (or changes during the operating period itself) somehow will implicate site operation as causal. We address that problem under Analytical Sampling (Section 4.3) and superficially in other sections. We have included all such problems under Design because there is no clearcut question being asked (i.e., the relevant report section depends on the question(s) being asked). Thus, the interested reader may need to review more than one section of the report, and substitute the particular parameter of interest (i.e., chemistry, porosity, air quality measures) because our examples are based mostly on radionuclides.

Page	Quotation	Statistical Implications
14	"The proposed rule requires that the applicant conduct a pre-operational monitoring program to provide baseline data on site characteristics." (The rest of the paragraph implies these same sampling locations should be used during operation and closure.)	Design.
15	"The staff anticipates that a minimum record of one year of site-specific meteorological data will be collected during site characterization."	Design.
16	"Parshall flumes and sediment samplers should be installed, where appropriate, to determine the rate of erosion of site soil deposits under varying gradients and ground conditions."	Design.
18	"The parameters which govern the rate, direction, and velocity of of soil moisture movement and ground-water flow should be measured. These parameters include hydraulic conductivity, head, hydraulic gradient, effective porosity, specific storage, saturated thickness, and compressibility. The site characterization studies should establish mean and median values, ranges of values, and spatial and temporal variability for these parameters for the significant hydrogeologic units. The data collection points for moisture content and head measurements (above and below the water table) should be appropriately located in three dimensions within the separate hydrogeologic units to establish the boundary conditions for modeling."	Design. Sampling for Modelling.
19	"The water-holding parameters for soil above the water table, such as specific retention field capacity, and wilting point should also be measured."	Design.
19	"The frequency of measurement should reflect the mode of recharge of the hydrogeologic unit, with more frequent (weekly) measurements in unconfined units responding to precipitation events and less frequent (monthly)	Design.

Page	Quotation	Statistical Implications
	measurements in confined units not directly responding to precipitation events."	
19	"The site characterization studies should also include collection of ground-water chemistry data. Parameters measured should include indicators of drinking water quality and the direction of the flow, areas of recharge, and age of the ground water in the flow system. Other parameters which may affect the solubility of waste leachates, such as by causing changes in oxidation states and/or precipitation from solution, should also be measured. Major inorganic constituents, dissolved oxygen, and other characteristic properties, such as temperature and pH, should be included in the site characterization studies. Major organic constituents, such as total organic carbon and chemical oxygen demand, may also be included. Age-dating techniques such as tritium, oxygen-18/oxygen-16 ratio, carbon-13/carbon-12 ratio, carbon-14, and chlorine 36, should be included as appropriate for sites where disposal will be below the water table and molecular diffusion must dominate the flow system."	Design. Analysis of Ratios. (Not addressed in this report.)
19-20	"The staff anticipates that ground-water chemistry will be measured, at least quarterly, for a minimum period of one year and will be compared to historic records, if available. In this manner, the applicant can estimate seasonal and long-term fluctuations of ground-water chemistry as a background for continued monitoring of key indicators of ground-water quality and leachate (radiological and non-radiological) migration."	Design.
21	"The second type consists of exploratory boreholes with detailed descriptions and laboratory testing of core samples. To the extent practicable, the boreholes should be located, drilled	Design.

Page	Quotation	Statistical Implications
	and sampled for use in the hydrogeologic and geomechanical investigations and the site ground-water monitoring program. The third type of investigation, surficial geophysical surveys, should provide subsurface information between boreholes. These surficial geophysical surveys, such as electrical resistivity or seismic reaction should define the vertical and horizontal extent of the site stratigraphic units, heterogeneities, such as sand lenses within the units, and the surface of the water table."	
21	"Lithologic investigations will range from lithologic descriptions based on visual observations of soil samples and exposures in test pits to laboratory tests designed to produce data on characteristics such as clay mineral composition, carbonate content, particle-size distribution, moisture content, porosity, permeability, bulk density, and ion-exchange capacity."	Design.
24	"Air quality parameters measured in the site characterization program should define the level of airborne radionuclides contributed by atmospheric fallout, natural radiation released from the soil, and agricultural activities such as spray application of fertilizers or insecticides. Suspended particulates should also be measured during the site characterization studies."	Design.
24	"The measurement of air quality parameters should provide at least a one-year record of site-specific information. These data will serve as a baseline for the site monitoring program, and the data collection points used during site characterization should be incorporated into the site monitoring program."	Design.

1.3 APPROACH AND FUTURE WORK

The original statement of work included the following seven tasks;

1. Survey of statistical and sampling needs for environmental monitoring of commercial low-level radioactive waste (LLW) disposal facilities,
2. Recommendations on sampling for inventory at commercial LLW sites,
3. Recommendations on sampling for spatial estimation of pattern at commercial LLW sites,
4. Recommendations on sampling for comparisons needed to analyze data from commercial LLW sites,
5. Recommendations on sampling for commercial LLW site modelling,
6. Recommendations on site characterization,
7. Recommendations on monitoring radioactivity in the environment at commercial LLW disposal sites.

The purpose of Task 1 is to develop guidance for determining the overall needs for sampling and statistical work in characterizing, surveying, and monitoring LLW sites. As such, it is essentially a "pilot project" for the overall program. Results from work specifically conducted under Task 1 are in Chapters 2.0 and 3.0. In addition, parts of Sections 4.1 to 4.5 also resulted from investigations conducted in Task 1. Substantial progress on Tasks 2-5 is reported in Sections 4.2 to 4.5. At this writing, we have just begun Tasks 6 and 7 and preliminary findings are in the Executive Summary while areas for future investigation are in Chapter 5.0. The completion of Tasks 1-5 (the investigations identified in Chapter 5.0) will result in reports outlining our recommendations (Tasks 6 and 7).

The emphasis for the further investigation needs some careful consideration with respect to likely future needs in the low-level waste management program. Over three hundred papers have been examined in drafting this report (these will be presented in a separate bibliography to be submitted to NRC during FY 1983). A number of those papers discuss sampling strategies, and we see little need to re-examine this methodology as such. What is essential, we believe, is to try to identify those approaches that are actually appropriate, and not merely another repetition of standard textbook formulations. Unfortunately, a sizable fraction of those recapitulations noted in the literature fail to recognize that textbook methodology rests on theory based on certain assumptions. Some of these assumptions simply do not hold in low-level waste management situations, so that this initial report had to be partially devoted to reconsideration of some fundamental ideas.

1.3. APPROACH AND FUTURE WORK

The original statement of work included the following seven tasks:

1. Survey of statistical and sampling needs for environmental monitoring of commercial low-level radioactive waste (LLW) disposal facilities.
2. Recommendations on sampling for inventory at commercial LLW sites.
3. Recommendations on sampling for spatial estimation of pattern at commercial LLW sites.
4. Recommendations on sampling for comparison needed to analyze data from commercial LLW sites.
5. Recommendations on sampling for commercial LLW site modeling.
6. Recommendations on site characterization.
7. Recommendations on monitoring radioactivity in the environment at commercial LLW disposal sites.

The purpose of Task 1 is to develop guidance for determining the overall needs for sampling and statistical work in characterizing, surveying, and monitoring LLW sites. As such, it is essentially a "pilot project" for the overall program. Results from work specifically conducted under Task 1 are in Chapters 2.0 and 3.0. In addition, parts of Sections 4.0 to 4.5 also resulted from investigations conducted in Task 1. Substantial progress on Tasks 2-5 is reported in Sections 4.5 to 4.9. As this writing, we have just begun Tasks 6 and 7 and preliminary findings are in the Executive Summary while plans for future investigation are in Chapter 5.0. The completion of Tasks 1-5 (the investigations identified in Chapter 2.0) will result in reports outlining our recommendations (Tasks 2 and 3).

The approach for the further investigation needs some careful consideration with respect to fully recognizing the low-level waste management program. Over the hundred papers have been examined in draft form. This report (which will be presented in a separate bibliography to be submitted to NRC during FY 1983). A number of these papers discuss sampling strategies, and we see little need to re-examine this methodology as such. What is essential, we believe, is to try to identify those approaches that are actually appropriate, and not merely another repetition of standard textbook formulations. Unfortunately, a substantial fraction of those recommendations noted in the literature fail to recognize that textbook methodology rests on theory based on certain assumptions. Some of these assumptions simply do not hold for low-level waste management situations, so that this initial report had to be particularly devoted to reconsideration of some fundamental ideas.

2.0 BACKGROUND

2.1 INTRODUCTION

Many scientists tend to think almost exclusively in terms of an experimental approach to both basic and applied problems. To search for a new fundamental particle predicted by theory, or to improve a fertilizer, one starts with an experimental design. Most such designs include several "treatments," and provide both controls and replication against temporal, spatial, and other factors that may be extraneous to the "main effects" of primary interest. The random assignment of treatments to relatively homogenous experimental material is the mainstay of such a design, and provides the essential basis for a statistical analysis of the results from experiments.

The ability to control experimental conditions (thus balancing out unanticipated, extraneous effects) is basic in modern science whether or not statistical methods are also used. In waste management, as in many other areas of practical importance, there may be no opportunity to design an experiment. Instead, the researcher must observe, through sampling, an uncontrolled and often complex and variable system. The circumstances are thus quite different from those regarded as basic for experimentation. The "treatments" depend on factors beyond the observer's control, and may, in fact, be such that they cannot be identified or measured reliably. There is no randomization in the choice of experimental units, since only a single site is usually involved.

There are, of course, situations in waste management in which controlled experiments can be conducted. For example, the effects and behavior of hazardous substances contained in wastes can be studied under laboratory conditions. These experiments are almost always small-scale investigations that must be "scaled up," often through a model, to field conditions. One would thus like to use the results of field samplings to assist in and check the scaling.

Study designs may be separated into three rather distinct categories:

1. Designed experiments (treatments are assigned randomly to selected experimental units);
2. Assessment of a time series (a relatively long series of observations is required); and
3. Sampling (taking controlled observations of a well-defined population).

In the waste management context, the experiments will usually be field or laboratory studies of some parameter of importance (e.g., rates of water percolation through different soils), while the time series data will largely come from monitoring (e.g., air samplers operated at disposal sites). Sampling is required in a wide range of circumstances. Examples of the three classes of designs are suggested in Figure 2.1.

FIGURE 2.1. Three classes of study designs. The first (1) is the traditional agricultural experimental design, while the second (2) refers to statistical approaches used to deal with long series of observations, and the third (3) considers planned sampling.

The general problem described above is one that is shared by other scientific disciplines, including astronomy, economics, ecology, and much of social science. In these fields and some others, we can usually only observe some phenomenon without controlling most conditions except the time and place at which the observations are made. One can, of course, define the population to be sampled in a way that makes the sampling easier. An example is provided by the NRC Regulatory Guide 4.5 (Measurements of Radionuclides in the Environment, May, 1974), where it is recommended that soil sampling follow the methodology used by the former Health and Safety Laboratory (HASL) to study worldwide fallout. In this scheme, sites are carefully selected to avoid possible disturbing influences (no runoff or other erosive forces, away from trees or buildings, uniform soil, etc.). Such an approach presumes that only a few samples are to be taken, and that the purpose is to estimate deposition from a very uniform source (such as worldwide fallout) over a substantial time period. In the low-level waste situation, it appears essential to consider that a number of other forces may be involved in radionuclide distribution, including such factors as:

1. Lateral transport by water (movement down a "wash," for example);
2. Penetration into the soil;
3. Gradual dispersion along ground surfaces, as the result of wind; and
4. Local accumulation due to topographic features or location of vegetation.

An appropriate sampling scheme will thus take into account the prospect of changes since deposition occurred, as well as the possibility of uneven distribution. In other words, sampling schemes suitable for world-wide fallout may be inappropriate for low-level waste sites. The HASL sites could be typified by golf courses and lawns where long-term accumulation of global fallout could be expected to take place with a minimum of disturbance. Sampling at low-level waste sites, on the other hand, will be most important in the event that disturbance actually takes place (see Figure 2.2).

Development of sampling methods for dealing with contaminants and pollutants has generally lacked the sharp focus provided by management needs in studies of natural resources. That focus has largely been provided by the need to find ways of efficient utilization and exploitation of resources, which translate readily into questions of "how much?" or "how many?" and, thus, permits direct application of sample survey methods. Several different objectives in sampling need to be considered with respect to low-level waste studies. One arrangement of objectives follows:

SAMPLING FOR WORLDWIDE FALLOUT

A POTENTIAL WASTE SITE SITUATION

FIGURE 2.2. A comparison of sampling for worldwide fallout and sampling at low-level waste sites. For worldwide fallout, undisturbed sampling sites were sought out (e.g., lawns and golf courses), but this is not useful at low-level waste sites where there may be high variability over a site due to such factors as lateral transport of radionuclides along preferred pathways (e.g., along a "wash" or local accumulation due to topography or vegetation).

1. Estimation of a material balance

The main goal of such a survey may be taken to be estimation of the total amount of some substances (e.g., radiocesium, plutonium) in the environs of a waste disposal site. If the primary objective is actually one of estimating the total amount present, the sampling procedures may be quite straightforward. However, in waste management the principal concern usually is not in determining quantities present, except in special circumstances.

2. Establishment of baselines for monitoring

Establishment of baseline levels in the environs of a low-level waste site is clearly important throughout the effective life of a site and in decommissioning and early custodial stages.

3. Assessment of consequences on human health

Transport, uptake, and retention of radionuclides are essential factors in considering potential consequences on human health due to nuclear wastes. However, the existing management guidance for low-level waste sites sets both on-site and off-site radioactivity levels well below any point where a concern for consequences on human health can be considered

in designing a sampling program. The program is instead normally directed at assuring that those standards are not exceeded.

4. Evaluation of system behavior

Movements of radioactivity from a waste site may occur by several routes, usually considered to include plant, animal, or human intrusions, erosion and ground-water movement. Possible pathways to humans are of critical concern, as is chemical form, and both require close attention. Sampling and statistical analyses must thus be concerned with assessment of these processes with respect to the overall behavior of the transport systems.

The issues, objectives, and needs faced in low-level waste management are evidently sufficiently diverse to require a range of sampling approaches. These will be described later in the report, after some other relevant aspects are discussed in Chapter 3.0.

2.2 MODELS

An unresolved issue is whether or not very large and complex models can be trusted to provide the kinds of projections and forecasts required for effective waste management. It may well be that much simpler representations of the same phenomena are just as accurate and just as useful. One of the advantages of "small" models is that they can be fit to data, using non-linear least-squares techniques. Possibly large models serve only to entertain and hopefully to educate the modeller.

The process of "sampling for modelling" is a particularly difficult one to deal with here. All we expect to attempt in the present preliminary study is to try to make an explicit statement of our perceptions of the prospects for utilizing some particular approaches. In preparing for the rulemaking promulgated in 10 CFR 61, Smith and Hawkins (1980) described an extensive NRC effort to establish an in-house capability in analytical modelling. They (ibid:37) reported that models will be used to "simulate the geological environment, ground and surface water pathways, air pathways, biotic and vegetation pathways, radionuclide transport, and the assessment of potential radiological and non-radiological impacts." It was also noted that "In some cases, site screening and gross evaluation types of models may be all that are required; whereas for other cases very detailed, highly accurate models may be needed to fully assess impacts." An extensive program of evaluations, categorization, and comparisons was also described.

In a project such as this, especially in its initial stages, we cannot deal with all of the statistical and sampling problems inherent in such a substantial effort. We must instead begin with some manageable portion of the overall complex of issues concerning modelling. We note also that there are questions about the utility of the larger, more complex models. For example: "Despite active research, the credibility of groundwater models for long-duration predictions of natural systems

is yet to be firmly established." Also, "considerable difficulties exist in handling the interaction of statistical quantities that are mutually dependent or correlated." (Narasimhan 1979:112).

Some difficulties in utilizing models for low-level waste management include:

1. Adequacy of sampling - - In many cases, insufficient data are collected to support any realistic modelling effort.
2. Lack of knowledge of relevant parameters and processes - - Too many of the models used infer both processes and parameter values from theory or other studies, without any assurance that the transfer is appropriate and sensible. Characterization of both sites and chemical forms of wastes are particularly important for development of realistic models.
3. The phenomenon of feed-back - - In biological systems, radioactive substances may be repeatedly recycled, leading to great complexity in the models. Body size is highly (inversely) correlated to uptake and retention of trace substances, further complicating assessment of individual species in a biological system.
4. Relative masses of biota and abiotic environment - - Biological vectors may be of substantial importance in potential transport of radionuclides to man, but constitute only a very small portion, by mass, of the overall environment. The great majority of radioactivity is thus confined to and transported by the abiotic system, so that the couplings to the biota must be carefully considered if the model is to be at all useful in predicting human exposures. That is, pathways to humans often cannot be predicted from knowledge of the physical system and "mass balance" considerations.

2.2.1 Projections Versus Forecasting

An overall concern about models is simply whether a given model has any substantial predictive value. A distinction needs to be made between projections of an observed trend and forecasting something that will happen in the future. Projections are usually based on some sort of regression analysis of historical data. Forecasts depend quite heavily on regression and time series methodology but also incorporate elements of judgement and theory that may serve to alter the projections (cf. Levenbach and Cleary 1981; Cleary and Levenbach 1982). In dealing with low-level waste repositories, there would appear to be more interest in forecasts, inasmuch as simple projections may not be very revealing, given that the repositories are satisfactorily managed. Most of the useful forecasts will likely take a "what if" or "scenario" point of view. Some examples are available in a study of biotic pathways (McKenzie et al. 1982). There are, of course, many more facets to the analysis of long-term forecasting, especially in reference to requirements for ultimate decommissioning of a waste repository (some recent experience is summarized by Phillips et al. 1980).

Our experience has largely been that the simpler models are most likely to be useful, often through regression analysis of the available data. Much more detailed simulations are largely valuable as a way to study the importance of the many possible deviations from the simple regressions assumed in a model. It should be noted that regression is here used to include both linear and non-linear models, and fitting models with non-linear least-squares may thus be required. There is, of course, a need to be more specific about what is regarded as a "simple" model. Lu (1978) describes a "simplified model" for underground migration of radionuclides at the West Valley site that illustrates very well the complexity inherent in attempting to construct a model from first principles. A great deal of further simplification is required if sampling strategies are to be considered, using actual field data.

Cherry et al. (1979) suggest some other issues worth considering here. They note that desirable subsurface waste storage sites are those in which there is evidence that rates of groundwater flow are very low, but that such sites have not received much attention by hydrogeologists because they are of little interest in normal practice. In non-arid areas such sites, when found near the surface and in readily excavated overburden, are largely in clay or silty clay material. In homogenous situations, the appropriate models may then be mostly concerned with diffusion processes. However, the presence of any kind of jointing or fracturing of the deposits necessarily makes a diffusion model meaningless for that zone. Hence, elaborate models based on diffusion may go widely astray of the desired predictions. We believe that this is one of the major pitfalls in use of models in low-level waste situations, and it may be noted that movement of radioactivity from several existing waste sites may be attributed to this cause.

2.2.2 Deterministic, Stochastic, and Simulation Models

Differences between deterministic, stochastic (probabilistic), and simulation models need to be considered. Many of the groundwater models are deterministic, based on systems of differential or partial differential equations with no reference to chance fluctuations. Complexity enters into these models mainly because the solutions very often must be numerical via a computer. Stochastic models attempt to reproduce various chance events by introducing such features as "noise" and probability mechanisms to govern events, and thus are often inherently complex in structure.

The rather elaborate schemes for spatial sampling (e.g., kriging) are based on stochastic models, which may also be very difficult to handle, leading again to solutions via computer. Simplifying assumptions and approximations are nonetheless necessary if moderate quantities of data are to be used in the analysis.

Simulation models can incorporate both deterministic and stochastic elements, and can be resorted to as a way to avoid some of the difficulties inherent in both kinds of models. Perhaps an even more compelling reason for resorting to simulation is inadequate data. When data are lacking, the simulation modeller can substitute any one of a number of

alternatives, usually in the form of some kind of "interpolative" function or by plucking a needed constant from some other circumstances. If the model is large and complex, users may not realize just how little the actual data at hand have to do with the model outcomes. We believe that this is often true in practice, especially when "user" groups do not participate in model construction and testing.

2.2.3 Classes of Models

Several classes of models may need to be considered in the kinds of studies contemplated here:

1. Linear differential equations with constant coefficients

These models constitute the generally used framework of "kinetic" models. However, the actual scope of such kinetic models is much wider, and this class of models is frequently unsatisfactory in assessing actual field data because some supposed constants are actually a function of time.

2. Linear differential equations with coefficients that are functions of time

In practice, the coefficients changing with time are usually "transfer" coefficients, coupling compartments in a system.

3. "Particle" models

In some situations, it is useful to consider individual particles of radioactivity, leading to stochastic models. Such models are not very useful for most of the circumstances considered here. However, it can be noted that the underlying actual processes may indeed have such a mechanism.

4. "Mass transfer" models

In some circumstances, it is useful to consider the history, over time, of individual "packets" of radioactivity.

5. Linear differential equations with coefficients as random variables

Models of this kind can be constructed by considering transfer coefficients as being randomly drawn from a frequency distribution of the possible values.

2.3 FREQUENCY DISTRIBUTIONS AND TRANSFORMATIONS

Frequency distributions may serve to provide a choice of a framework for statistical analysis of the data. There is often a need to examine particular sets of data to determine whether the methodology proposed is adequate. Analysis of data on transuranic elements may be particularly difficult, due to the very high sample-to-sample variability.

Administrators will occasionally remark that a result only needs to be "in the ballpark" to be useful. With coefficients of variation on the order of unity, as frequently observed for plutonium, it can be shown that such a "ballpark" is measured in orders of magnitude. Under the circumstances of such high variability, it is also true that the models employed must be stochastic, not deterministic.

A key assumption in most standard, textbook methodology is that of normal and additive errors. In working with radionuclides, this assumption is simply wrong in a majority of cases because data on radionuclides show skewed frequency distributions and relatively constant coefficients of variation (Eberhardt et al. 1976). We thus suggest that one of the first things to look for in evaluating a recommended sampling scheme is whether the proposed model has an "additive error term."

As soon as it is recognized that additive errors may not be appropriate, many of the existing recommendations on sampling must be reconsidered. The alternatives are not always very palatable, but can be lived with quite satisfactorily if one is willing to utilize some approximations. These are much more acceptable if it is appreciated that field sampling differs from experimentation in many ways. The simplest alternative is to assume a multiplicative model and lognormality. Making a logarithmic transformation then lets one proceed under the existing theory. However, several features have to be considered:

1. The answers to basic questions may be different. Formulae taken from standard textbooks no longer can be used directly.
2. The assumption of multiplicative effects and lognormality is itself probably not exact in most circumstances. Two refuges then exist. One is that there is a law of large numbers ("central limit theorem") appropriate in the multiplicative scale. Another is to do simulations and other studies to examine the consequences of a wrong choice of frequency distribution. Various aspects of these problems are discussed in some detail by Eberhardt et al. (1976), White (1978), and White and Hakonson (1979).
3. Many inferential problems can be avoided by staying on the logarithmic scale, i.e., discussing things in terms of logarithms. Since logarithmic shovels are not available, certain situations force "transforming back," and these offer difficulties that are discussed below. That is, any kind of "clean-up" operation will most likely require estimates on the original (untransformed) scale of measurement.

Administrators will occasionally remark that a result only needs to be "in the ballpark" to be useful. With coefficients of variation on the order of unity, as frequently observed for planning, it can be shown that such a "ballpark" is measured in orders of magnitude. Under the circumstances of such high variability, it is also true that the models employed must be stochastic, not deterministic.

A key assumption in most standard textbook methodology is that of normal and additive errors. In working with radioisotopes, this assumption is simply wrong in a majority of cases because data on radioisotopes show skewed frequency distributions and relatively constant coefficients of variation (Eberhardt et al., 1976). We thus suggest that one of the first things to look for in evaluating a recommended sampling scheme is whether the proposed model has an "additive error term."

As soon as it is recognized that additive errors may not be appropriate, many of the existing recommendations on sampling must be reconsidered. The alternatives are not always very palatable, but can be fixed with quite satisfactory if one is willing to utilize some approximations. There are much more acceptable if it is appreciated that fixed sampling differs from experimentation in many ways. The simplest alternative is to assume a multiplicative model and lognormality. Making a logarithmic transformation then lets one proceed under the existing theory. However, several features have to be considered:

1. The answers to basic questions may be different. Formulas taken from standard textbooks no longer can be used directly.
2. The assumption of multiplicative effects and lognormality is itself probably not exact in most circumstances. Two regimes then exist. One is that there is a law of large numbers ("central limit theorem") appropriate in the multiplicative scale. Another is to do simulations and other studies to examine the consequences of a wrong choice of frequency distribution. Various aspects of these problems are discussed in some detail by Eberhardt et al. (1976), White (1978), and White and Robinson (1979).
3. Many inferential problems can be avoided by staying on the logarithmic scale, i.e., discussing things in terms of logarithms. Since logarithmic answers are not available, certain situations force "transforming back," and these offer difficulties that are discussed below. That is, any kind of "clean-up" operation will distort likely negative estimates on the original (untransformed) scale of measurement.

3.0 SAMPLING UNITS

3.1 INTRODUCTION

An essential, preliminary step in designing a sampling program is deciding what sampling units will be used; that is, determining what should actually be measured or collected in the field. Sometimes the units are readily defined, since they are "natural" units (e.g., trees, people). However, if, for example, soil samples are to be taken, some care in planning can save much time and money. We will briefly address compositing or "pooling" samples as a way to improve sampling efficiency, point out some common mistakes, and suggest areas for future research. In addition, we point out some additional research needed, because less-than detectable concentrations are often reported.

3.2 COMPOSITING

In this section we discuss compositing of field samples (i.e., air, water, biota, soils, etc.). A working definition of a composite sample obtained from LLW sites might be the mass of material resulting from pooling several individual samples together before radioanalysis is done. Compositing samples from LLW sites will become attractive when the cost of a single analysis is large relative to costs of collecting, pooling, and adequately mixing samples. Two purposes of compositing are illustrated in Figure 3.1. In the first case, the desired outcome is an estimate of average concentration of a radionuclide or chemical in some biotic or abiotic component which might move off-site [10 CFR 61; Section 61.80(i)]. In contrast, another purpose might be detection of on-site spills, areas of unacceptably high radioactivity or possible radionuclide migration during routine site operation [10 CFR 61; Section 61.53(i)]. The desired outcome in the first case is an estimate of concentration and an appropriate estimate of variance. When compositing for detection, the desired outcome is a statement of the probability that none of the samples making up the composite contained more than some amount (for example, 1 nCi/g) of radioactivity and a confidence limit about the probability estimate. Both compositing purposes are discussed below.

Compositing samples for either purpose should only be done after careful consideration of possible problems. For instance, Eberhardt (1975) shows that for lognormally distributed data (almost always a tenable assumption for radionuclides in environmental samples), the mean value can be over- or underestimated by 1.65 using composited samples compared to individual sample analyses. He also advocates that experiments to estimate components of variance be conducted in cases where subsamples of composites are analyzed. Additional admonitions are in Gilbert (1978), who also reviewed three relevant papers (Brown and Fisher 1972, Kussmaul and Anderson 1967, and Rohde 1976). The focus of Gilbert's report was on particulate transuranic radionuclides where large analytic, counting and spatial variance components are not unusual. While relevant to existing LLW sites and perhaps some new sites, it appears that nonparticulate radionuclides, and thus smaller variance components, are likely to be most abundant at new LLW sites.

FIGURE 3.1. Compositing of samples can be used as a means of keeping the costs of analyses down. In many situations it is unnecessary and prohibitively expensive to analyze every sample. In the example shown, the analytical program might be designed to address one of two objectives -- estimation or detection.

3.3 COMPOSITING FOR ESTIMATION

The customary approach to evaluating a compositing sampling procedure is to construct a random effects linear model and use this model to derive the estimator for the variance of the mean (Duncan 1962). Brown and Fisher (1972), in studies of compositing bales of wool, introduced a model based on properties of randomly weighted averages. We note that Brown and Fisher (1972) discuss the case of one composite (samples from several different bales are the components of the composite), but do not include analytical error, or variability within the components that make up the composite. In contrast, Rohde (1976) presents a situation where the composite is made up of components from a lake. He derived a variance of the mean, but again allowed no analytic error and no within-component error.

FIGURE 3.2. Hypothetical example of compositing for estimation. Four random site soil or water samples (n) are combined (large boxes) and mixed. Four (r) of these composites are shown. These are subsampled (s times) and each subsample is split into t units where all of each subsample is used. Each unit (t) is analyzed separately. The dark shaded boxes depict the usual scheme used for soil ($r=s=t=1$) and water ($r=1$). If water cannot be evaporated (to reduce volume) sometimes a subsample (dashed box) is taken for analysis. The usual scheme (depicted by dark shaded boxes; see text) allows no estimate of variability of the original composite.

In a more recent paper, Elder, Thompson and Myers (1980) extended the previous work to the general case, where each of r composites is made up of n components (samples). Randomly selected subsamples (s) of each composite are subjected to t analyses on each subsample (Figure 3.2). The subsample is of a size that is all used in t chemical analyses. The model can be extended to the case where only part of the subsample is used by adding another subsampling stage. Their expression for the variance of the mean incorporates and accounts for all sources of variation for the situation where components are obtained from separate entities (the bales of wool in Brown and Fisher, 1972) or a bulk sample (the lake in Rohde, 1976).

We note here that in the absence of estimates of the respective mean squares between composites, between subsamples within composites, and between tests within subsamples (which depend on the compositing scheme used), that no statistically valid variance of the mean of the composite(s) can be obtained. Even so, it appears that the incomplete scheme for analyses of soil and water composites shown by shaded dark boxes and arrows in Figure 3.2 is usually employed. In general, an entire water sample may be subjected to radioanalysis (i.e., evaporated if need be) while only subsamples of soil may be evaluated (Figure 3.2). No variance estimate of the composite mean can be obtained based on the usual methods depicted in the figure (shaded dark boxes). A complete analysis similar to that illustrated for four soil composites would be needed to obtain estimates of the needed mean squares and/or variance components. Only a demonstration that the respective variances remain constant can justify using the usual scheme (one possible quality control is given at the end of this section).

The equations for the variance of the mean derived by Elder, Thompson and Myers (1980) are complex, but they show that the resulting estimate is equivalent to the usual mean square between composites (with appropriate divisor) obtained from a balanced two-fold nested classification. The difference lies in the variance components that make up the expected value of the mean square between composites. An evaluation shows that changing the number of components that make up a composite will affect the variance in a way not consistent with the usual balanced two-fold nested model. This is a consequence of the fact that larger composites are harder to "perfectly" mix and because larger composites contained smaller fractions of each component. Because of the difficulties in batch mixing (n) components, it might be worthwhile to compare results obtained using batch mixing with those from the "splitting" procedure outlined in Nyhan et al. (1981).

Thus, in order to obtain a correct variance estimate, an additional set of analyses (i.e., at least two of the soil composites depicted in Figure 3.2) must be conducted for each change in the number of components that make up a single composite. If the number of components in each composite remains constant, the optimal allocation of r , s , and t can be calculated (considering cost and effect on the variance of mean) in a straightforward way.

Thus, from our preliminary evaluation it appears that a practical procedure useful for site characterization at LLW sites (after variance components have been estimated) might be to run two subsamples of a single composite routinely, but periodically run two composites with a single subsample of each as a cross check on variances. This first strategy would be cheaper, because forming composites is more expensive than taking two subsamples (two radioanalyses are needed in either case). Clearly, actual site data and experience will be needed to substantiate the procedure outlined above. We note, too, that spatial heterogeneity over the site might lead to compositing within strata or some other sampling scheme.

3.4 COMPOSITING FOR DETECTION

This compositing technique is sometimes called group testing and was apparently developed during World War II to reduce the costs of blood testing in a program to weed out all syphilitic men called up for induction (Dorfman 1943). Rather than analyze each individual blood sample, Dorfman (1943) devised a much more cost-effective method based on pooled or composited samples. Hwang (1976) extended some of the ideas for presence-absence (binomial) group testing using nested designs. Recently, however, these methods have been suggested for evaluating toxic chemical waste sites (Schaeffer, Kerster and Janardan 1982). Unfortunately, these authors suggest the methodology, but offer little guidance for its actual use; nor do they outline the research that would be needed prior to use at an actual waste site.

Some possible scenarios for detecting spills at LLW sites are illustrated in Figure 3.3. The extension to the problem of migration may not be straightforward and will be investigated in future work on this project. All sixteen soil samples are composited in Scenario 1, a random selection of four samples in each composite is depicted in Scenario 2, while in Scenario 3, the compositing is done within four locations (A through D, possibly based on some prior knowledge). The apparent spills or hot spots are shown as the open circles in the figure. For purposes of illustration, we assume that when one hot spot is composited with three background samples, the radioactivity will not be detected in the subsample. In Scenario 1, detection of spills would actually depend on the level of radioactivity, the dilution by twelve background soil samples, the ability to mix 16 samples uniformly, and the subsample size. Under the assumption above, the contamination would go undetected. A spill probably would be detected under Scenario 2 (unless only 1 hot sample happened to be selected for each of the four composites), but the location would have to be determined by an additional analysis of the four individual components. Site C would be identified as contaminated under Scenario 3, but Site B could be missed. Many other possible scenarios could be constructed. Thus it appears that the development of a generalized strategy is needed.

At this early point in our research, we have defined the following statistical issues which must be resolved before group testing can be directly applied to the assessment of LLW sites:

1. The influence of a dilution effect when a contaminated sample is combined with several uncontaminated samples;
2. The effect of mixtures of non-homogeneous materials on the resulting concentration levels in composited samples;
3. The interpretation of test results when numerous radionuclides are the target of an analysis; and
4. Determination of optimal grouping strategies when the amount of contamination is unknown.

FIGURE 3.3. Hypothetical example of compositing for detection. For purposes of illustration we assume a spill site soil sample (open circles) cannot be detected when composited in a 3 to 1 ratio with background samples. Many other spill and compositing scenarios are possible depending on how the area was divided (we selected areas A through D arbitrarily) and the number of samples composited and their location relative to a spill and other factors.

If these issues associated with the application and interpretation of group testing methodology were addressed, the result could be group testing procedures tailored to spill detection and/or radionuclide migration at LLW sites.

Additional research is suggested by Gilbert (1982), who reviewed systematic sampling designs useful in detecting surface and subsurface zones of contamination, or hot spots. Computer simulation models are being developed to test alternate sampling strategies at contrived LLW sites under various scenarios that call for the spatial dispersion of

radionuclides. Scenarios could vary from a point source of contamination, with the objective of defining boundary locations, to multiple point sources, with the objective of locating all major spills. Simulation studies incorporating group testing strategies could be performed for these scenarios, with the option of assessing how prior information (available at some sites) could be used to improve field sampling designs. Currently, little is known about how prior knowledge of LLW site history can be efficiently used to optimally design field inspections. Less is known about how a field design can be structured to minimize the consequences when the prior source(s) of information about contamination is erroneous. Research designed to provide specific guidance on the use of group testing and prior knowledge in the design of sampling programs at LLW sites appears to be needed.

To demonstrate the applicability of group testing procedures for compositing samples and improving the reliability of field surveys, the techniques developed should be implemented at an existing LLW site. One result of a field validation study would be a set of guidelines for the application of sampling strategies which have been evaluated by means of computer analysis and actual field testing.

Data from the radioanalysis of soil samples and field sampling costs could be used to determine the efficiency of the proposed procedures and their limitations relative to other practices likely to be used at LLW sites.

3.5 OPTIMAL VOLUME OR WEIGHT FOR SAMPLES

Another factor we have not considered in detail in this report is the optimal volume or weight of field samples (components) used to make up composites or the subsample size taken from the composites. Nyhan et al. (1981) show that optimal soil volume for Cesium-137 depends on how the radionuclide was dispersed and its bulk density. Doctor and Gilbert (1978), (using a single composite soil sample containing Americium-241) found a reduction in the coefficient of variation from 79% to 9% when 1-g and 100-g subsamples were compared. It appears from these studies that the largest possible sample volume or weight should be used. However, as noted above, this will affect the ability to mix the composite. Nyhan et al. (1981) illustrate the influence of mixing 12,500 cm³ soil samples on Cesium-137 concentration, and Gilbert (1978) presents various techniques for estimating the "mixing variance." We plan to provide additional guidance during future research on this topic.

3.6 CENSORED DATA SETS

Samples of soil, air and water have been routinely collected around nuclear facilities as part of monitoring programs. Limitations of prior data sets, including existing LLW sites (i.e., Nielson, Wogman and Kirby 1980 for the Beatty, Nevada site, and Commonwealth of Kentucky, 1982, for the Maxey Flats site), include the existence of "less than detectable" concentrations, unreported concentrations because they were below the detection limit, and sometimes negative values for radionuclide con-

centrations in soil, air, and water. Negative values indicate background counts are higher than those from a sample and are the easiest to evaluate statistically (Gilbert and Kinnison 1981). Data sets with these characteristics are termed left-censored. When the objective is to estimate the mean and variance for the biotic or abiotic component under study and the data are left-censored, serious complications arise (Gilbert and Kinnison 1981). While Gilbert and Kinnison (1981) present some useful methods, they do not discuss statistical needs appropriate for the statistical comparison of two censored data sets. Such a circumstance will arise at LLW sites when the soil concentration in one area is compared to another and both data sets contain replicate (Section 4.3.1) observations that are reported as less than detectable.

Radionuclide concentrations in environmental samples are generally considered to be left-censored because net concentrations "below" detection limits are reported as "less than detectable" values or sometimes as indicator zeros (zeros indicating concentrations below detection limits). We note that censored distributions are not synonymous with truncated data sets. In the former case, the number of values above and below the detection limit is known; in the truncated case, they are unknown. In contrast to environmental contamination research, survival studies of different populations in biomedical and industrial contexts (animals, humans, and light bulbs) usually result in right-censored data sets (Latta 1981; see Kane 1979 for an environmental example) because studies are terminated before the full "lifetime" of the study object. While there are dissimilarities in end points (i.e., a failure time compared to a radioactive count), the biomedical and industrial literature has apparently not been reviewed for applicability to environmental needs. This appears to be a fruitful line of confirmatory research. Gilbert and Kinnison (1981) also note that when the arithmetic mean of untransformed but lognormal and uncensored data is computed, the loss in efficiency is only 10% (for coefficients of variation up to 1.2). However, the loss in efficiency for estimating the variance is substantial (about 20% compared to using the log transform). Simulation studies designed to study the severity of censoring on the respective efficiencies is another aspect of research needed for LLW sites, since routine results are often summarized and reported on the arithmetic scale.

4.0 SAMPLING METHODS

This chapter of the report deals with the main subject of our research, sampling methods. In Chapters 2.0 and 3.0, we discussed some preliminary, but basic, matters that need to be considered in sampling programs. Here, we look in more detail at four broad classes of sampling (descriptive, analytical, pattern and sampling for modelling). These are introduced and briefly described in the next section (4.1). A more detailed analysis of each class of sampling is in Sections 4.2 through 4.4.

4.1 INTRODUCTION

Many users of sampling schemes assume that there is just one "correct" answer to the question, "How shall I sample?" In reality, a wide range of quite different answers can be arrived at by postulating different objectives for the sampling program, under a given set of assumptions about the circumstances in which the sampling will be carried out. Since we often cannot be sure of the correct choice of assumptions, the range of relevant answers can be further compounded.

In many circumstances, the four classes of sampling may usefully be applied sequentially. Initial surveys may be concentrated on an analytical approach, to determine whether there are substantial differences in subregions of a site. These differences, if found, then need to be assessed in more detail, using sampling for pattern.

Evaluation of the patterns thus found will then facilitate (and make cost-effective) any sampling for inventory that is required. In many instances, preliminary attention to sampling for modelling is needed inasmuch as subsurface features and ground-water flows will have to be approached from a modelling standpoint.

A somewhat over-simplified illustration was presented by Eberhardt and Gilbert (1980:177) to emphasize differences in three of the four classes of sampling discussed here. (The fourth class, sampling for modelling, did not seem relevant in this example). The appropriate allocation of 100 surface soil samples over four strata was approximately as follows:

Sampling for:

<u>Stratum</u>	<u>Inventory</u>	<u>Pattern</u>	<u>Comparison</u>
1	32	92	25
2	21	6	25
3	11	1	25
4	36	1	25
	100	100	100

If sampling is at all expensive, differences of this magnitude ought to suggest that an appreciable effort in designing the sampling program is worthwhile. That is, if the allocation had been based on the "inventory" approach, the results would have been very inefficient if they were used for depicting a pattern or making comparisons. In the example above, roughly two-thirds of the expenditures made for "inventory" purposes would be wasted if the main goal were sampling for pattern. The importance of the assumptions made about the circumstances in which sampling will be done can be illustrated by noting that the frequently-employed assumption of a common sampling variance in the four strata results in the allocation for "inventory" becoming essentially the same as that given for "sampling for pattern," which is quite a substantial change. In the case studied by Gilbert and Eberhardt (1980), the strata variances were not the same.

Most of the balance of this report will be devoted to making rather more explicit statements about the nature of the four classes of sampling. An initial description of the different classes follows. Figure 4.1 attempts to show how the four methods differ, with respect to waste management problems, in a generalized way. If wastes are lost from a trench, one may wish to determine the quantities lost through survey sampling (1), analyze differences between regions of the area (2), describe the pattern of spatial distribution (3), or obtain data efficiently for a model of distribution by depth (4), as shown in the figure.

4.1.1 Descriptive (Survey) Sampling

The classical sampling methodology seeks to describe a population of some kind by estimating totals or means from samples. Most of the texts with "sampling" in the title are devoted to descriptive sampling. In a series of studies of radioactivity at the Nevada Test Site, the term "inventory" was frequently attached to this kind of approach. It is, however, a poor term to use, since an inventory is usually perceived as providing a complete count of items in stock. If the tally is complete, presumably there is also a record of the location of each item. A useful discussion of sampling techniques for such a circumstance (inventorying special nuclear materials) is given by Piepel and Brouns (1982).

In the context of nuclear waste studies, it is usually impossible to make an "inventory" of the small quantities of radioactivity of interest (those outside of the repository proper), and we believe that it is important to separate the concept of estimating a total or a mean from that of locating the individual components of that total. At LLW sites, buried inventories (locations and amounts) are accomplished by record keeping. If any redistribution of the material occurs however, the estimation of quantities may then become an important objective.

4.1.2 Sampling for Comparisons (Analytical Sampling)

Over much of its history, survey sampling was largely considered in terms of estimation of a mean or total. More recently it has been appreciated that much of the data may also be utilized in an analytic

FIGURE 4.1. Four classes of sampling at a hypothetical low-level waste shallow land burial facility. In the inventory case (1), interest is mainly on determining a total, as suggested by integrating over the region of dissemination of the material. For comparisons (2), one wants to determine if, for example, radionuclides have moved further in one direction than in another. Patterns may, however, be represented by a contour map (3), while data to fit a mathematical model might be desired, such as fitting an exponential curve to concentration with depth (4).

mode, i.e., what can be said about differences between subpopulations or areas? In the LLW context, many of the questions about base line and monitoring are of this nature -- has there been a statistically significant change? The relevant methodology may thus be that of classical statistics -- hypothesis testing and the like. It is essential to appreciate that the inferences derived from such an analysis may be quite tricky because a sampling design usually cannot approximate an experimental design. Also, as suggested in Section 2.3, the usual model for most statistical analyses (normal additive errors) is inappropriate for work with radionuclides. The frequent presence of samples "below detection limits" poses another important issue in choice of methods (Section 3.6).

4.1.3 Sampling for Modelling

Almost any effort to assess a defined process involves use of a model of some kind. Frequently the parameters of such a model need to be estimated from observations made on the process. An appropriate choice of times or places for making these observations (sampling) may make for substantial improvements in estimating the desired parameters. Most of the applications thus far have been in industrial research and development. Two applications in LLW management may be suggested as examples. One is determining sampling times for monitoring biota in the event of the loss of significant amounts of radioactivity from a repository. A second example is in determining "profiles" of radioactivity with depth in soils or with distance away from a source. The sampling scheme ordinarily used (evenly spaced increments) will be quite inefficient because much of the sampling effort is then located at the wrong depths (Section 4.4.1).

4.1.4 Sampling for Pattern

When any movement of radioactivity from a LLW site is either experienced or anticipated, a major interest is likely to be in the pattern of concentrations in space and time. If a specific mathematical formulation for the process governing movement is available, the methods of "sampling for modelling" (Section 4.4) may be appropriate. When a somewhat more general appraisal is needed, the methods developed in "geostatistics" (Section 2.5.6) may be preferred.

4.2 DESCRIPTIVE (SURVEY) SAMPLING

4.2.1 Introduction

Although there is a well-known foundation for this class of sampling, with perhaps a dozen textbooks now on the scene, that fact does not guarantee that the advice available will be taken or will be appropriate in low-level waste management. Both first-hand observation and perusal of the literature suggest a need for more attention to elementary survey sampling in low-level waste management studies.

In many of the applications, the main need is to choose among a few well-defined sampling schemes. In this section, it is assumed that the

investigator will be able to decide how the sampling will be done. In some circumstances this is not possible, since the choice will already have been determined for some other purpose (e.g., drilling wells for domestic water). Issues concerned with those circumstances are discussed in Section 4.5 on sampling for pattern. Most textbooks (and many statisticians) recommend random sampling, largely on the basis that most sampling theory is based on random selection of sample locations. However, systematic sampling has two possible advantages. One is simply that it is easier to do in the field, and thus less costly. The other depends on the existence of small-scale pattern. If any two nearby locations tend to give similar results, then it may be advisable to space the sample locations as far apart as possible, and hence to use systematic sampling. Quite a lot of complications and complexity are contained in this simple-seeming recommendation, and should be appreciated before a decision is made. One first should be sure that nearby samples are indeed correlated (requiring methods for testing for spatial correlation), and the investigator should also be willing and able to accept the fact that he will most likely not get as good an estimate of the variance from systematic sampling. We will discuss this issue further below.

Another major consideration in selecting a sampling scheme is that of efficiency -- how can we get the best results for a given cost? Choices among various possible sampling schemes will largely depend on this matter of cost-effectiveness. In our experience, the main choices have been between stratified sampling and double-sampling (in which two measurement methods are used, one expensive and accurate, the other less accurate but inexpensive). These topics provide the basis for two further subsections in this section.

Another issue needing attention in sampling at low-level waste sites is that of repeated sampling in time. Often repeated sampling is done as part of a routine monitoring scheme, and most monitoring is carried out by taking the same kind of sample at a given place and on a regular schedule. A common example is the operation of an air sampler. The problem with such an approach is that if a poor choice of sampling locations is made initially, that choice is then perpetuated. Furthermore, if the samples are always taken at the same place and on a fixed time schedule, there may be no way to determine whether or not a poor choice was made in the first place. Hence, a fixed schedule on fixed locations may be unsatisfactory, and it is worthwhile to consider some alternatives, based on different sampling schedules in time and space.

4.2.2 Systematic Sampling

Systematic sampling is a topic on which much research has been done, and for which a large literature exists. Since the issues concerning when, where, and how to use systematic sampling are not resolved, it will be necessary to examine the topic in more detail than is possible in this preliminary report. For the present, we will only attempt to provide an impression of the issues and the coverage available in survey sampling references. Most authorities recommend that, if systematic sampling is to be used, it should be accomplished with a "random start." That is,

one first determines sample size, and thus sampling interval (which will involve taking every k th individual or location in a list or on a map, etc.), and then draws a random number to locate the first individual. From that point on, every k th individual falls in the sample.

The major hazard in systematic sampling is that there may be some kind of periodicity in the population, i.e., that some sort of regularity exists in the variation encountered from unit to unit. A well-known example is the kind of regularity imposed by human activities, e.g., roads, fence-rows, etc. tend to be at equally spaced intervals. If this is true, then a systematic sample may yield biased results. Much the same problem could be anticipated in a low-level waste context if, for example, trench markers were not present, and a search for trenches was made with a systematic sample. With the wrong interval, one could miss the trenches entirely, or hit them with every probe.

The next most important issue is that variance estimates derived from systematic samples are of uncertain validity, and often are overestimates, at least in the circumstances where systematic sampling is most useful, i.e., when there is a serial correlation i.e., adjacent units tend to be more alike than those some distance apart. If this is true, then it is clearly desirable to avoid including adjacent (or nearly so) units in the sample, since they tend to repeat the same information. One might then take uniformly spaced (systematic) samples to avoid this situation, which seems to happen more often in random sampling than might intuitively be expected. A "catch" in this logic, however, is that similarity in adjacent samples is exactly what would be expected if there is some underlying periodicity in the population. Hence the concerns about systematic sampling cannot be easily resolved. Of course, the correlation between adjacent samples can also come from a regular trend throughout the population (e.g., generally increasing or decreasing values of the variate of chief interest). In this case, the systematic sample gives a better estimate than a random sample, but the usual variance estimate is biased. In many circumstances, this can be avoided by using a number of systematic samples, each with a random start, and computing the variance from means or totals for these individual "cluster" samples. It may also be necessary to introduce "end corrections" when there is a trend in the population.

Cochran (1977:212) remarks that there are circumstances in which the variance of a systematic sample may increase as a larger sample is taken, which is, he notes... "a startling departure from good behavior." Cochran's summary on systematic sampling provides an excellent brief account of situations in which systematic sampling can be recommended:

1. Where the ordering of the population is essentially random or contains at most a mild stratification. Here systematic sampling is used for convenience, with little expectation of a gain in precision. Sample estimates of error that are reasonably unbiased are available.
2. Where a stratification with numerous strata is employed and an independent systematic sample is drawn from each stratum. The

effects of hidden periodicities tend to cancel out in this situation, and an estimate of error that is known to be an overestimate can be obtained.

3. For subsampling cluster units. In this case an unbiased or almost unbiased estimate of the sampling error can be obtained in most practical situations.
4. For sampling populations with variation of a continuous type, provided that an estimate of the sampling error is not regularly required. If a series of surveys of this type is being made, an occasional check on the sampling errors may be sufficient (by taking supplementary observations).

Hansen, Hurwitz, and Madow (1953:504) recommended that:

"In practice it is safe to use systematic sampling only when one is sufficiently acquainted with the data to be able to demonstrate that periodicities do not exist, or that the interval between the elements of the sample is not a multiple or submultiple of the period. The risk with systematic sampling from a population with periodicities is particularly serious because the sample itself may give no indication of the periodicity."

Kish(1965:120) suggested:

"Because ordinarily we cannot be absolutely certain of having avoided all danger, some statisticians prefer to avoid systematic sampling altogether. I, and many others, are cautious about the pitfalls of systematic selection, but take advantage of its convenience and simplicity in many practical selection situations. In most practical situations after investigating, we can dismiss the dangers both of a monotonic trend and of a periodic fluctuation coinciding with the selection interval."

Yates (1981:43) says:

"Systematic sampling is entirely unsuited to material which has periodic features, but apart from this will generally provide a satisfactory method of area sampling. It has the advantage over stratified random sampling from blocks that the location of the sampling units is simpler and the results obtained provide rather better material for the construction of maps, etc. As in systematic sampling from lists, however, the responsibility for the judgement that the material is such that systematic sampling will give satisfactory results rests with the investigator."

4.2.3 Cost-Effective Survey Sampling Methods

Two major reasons for using a sampling design are to reduce costs and avoid biases. In various circumstances it will not be possible to

obtain reliable results from a simple random sample. Usually this is a consequence of either record-keeping, or the physical circumstances in which the sampling must be done. If the material to be sampled has to be located through records, then it is essential to take into account the structure of the records in designing a sampling scheme. If plants are to be sampled (e.g., trees for tritium at Maxey Flats) it may not be feasible to take a simple random sample, since this would entail assigning a number to every tree in the population of interest.

Many opportunities exist for reducing costs through effective sampling designs. Most such opportunities depend on the fact that the investigator has some advance knowledge of special features of the material to be sampled. In low-level waste management contexts, these will often take the form of a general knowledge that some event of concern is more likely to occur in particular circumstances, places, or times (or that an event has already occurred in a given locale, such as a spill or other loss of radioactivity from confinement).

The main categories of special sampling techniques are as follows:

1. Stratified sampling. Some advance knowledge, or a preliminary survey, permits one to classify all possible sample units into several classes (strata) that differ with respect to the variable to be measured (or are expected to differ when the sampling is done). Almost always the material in each such class (stratum) will be less variable than the population of units as a whole. Hence, a weighted combination of samples taken independently from the several strata will yield a smaller variance than would be obtained from a simple random sample. Since stratified sampling is one of the most effective techniques for the purposes of this report, further discussion appears in a separate section below.
2. Ratio and regression estimation. In some circumstances, the investigator will have data on some auxiliary measurement that can be used to construct an improved estimate of the variable of interest. In many of the survey sampling uses of this approach, the auxiliary measurement is available on every unit in the population, and a sample is taken to measure the variable of interest (and to record the value of the auxiliary variable for those units that fall in the sample). In the low-level waste management context, this might be the case if it is necessary to sample some of the incoming waste containers, since the contents will be documented, including a measurement of radioactivity. However, the method will probably be of more interest in terms of the situation where two kinds of measurements can be taken on sampling units. A common example is soil sampling, where an accurate measurement calls for a laboratory determination, perhaps using wet chemistry and entailing a substantial cost. Often a field measurement instrument can be used at a much lower cost to obtain a less accurate estimate of the sample (and sometimes of an associated radionuclide, e.g., americium associated with plutonium in the samples). Usually the field instrument will not be used to survey all possible samples, but it can be used to survey a much larger sample than can be handled in the

laboratory. Since this, too, is a technique of wide applicability, it will be described in more detail below, under the heading of double-sampling.

3. Cluster sampling and subsampling. Sampling units are often made up of aggregates or clusters of some sort. Occasionally the clusters are all of the same size, i.e., each will have the same number of subunits. More often, there will not be the same number of subunits in each cluster. Sometimes the clusters are natural in origin, but they may also be defined in the process of sampling. In many cases, it is not possible to avoid cluster sampling, since it may not be feasible either to define or locate the individual units in advance, so that the sampling has to be done as a two--stage or multi-stage process. For example, an investigator may wish to assess radioactivity in tree leaves, using the individual leaf as a sampling unit. The units will obviously come in clusters, and a random sample cannot be selected in advance. There are, of course, other ways to approach sampling leaves. Compositing (Section 3.2) may usefully be considered as a cluster sampling process. One randomly selects units of area, subsamples these, and then composites the subsamples of each unit of area. The areas are then primary sampling units, which may then be subsampled in various ways. There is a variety of methods for situations of this kind, which we will not attempt to discuss in detail here. The important consideration in this preliminary report is that cluster sampling and subsampling can be used to handle situations that cannot be dealt with in any other way, and can be very cost-effective.

4.2.4 Further Details on Stratified Sampling

As suggested above, the basic idea of stratified sampling is to assign potential sampling elements to strata in such a way that the elements in each stratum are as much alike as possible. Sometimes this can best be done simply by advance knowledge of the area and material to be sampled, and in other circumstances it may be possible to do a quick preliminary survey of the area to set up strata. One of the simplest examples is the use of radiation detection survey instruments to set up strata for later sampling in which the analyses may be done by wet chemistry methods, or other fairly expensive analytical procedures.

The basic idea underlying stratified sampling can readily be appreciated by recalling that strata are normally layers of relatively homogenous material. Very likely, stratified sampling is one of the most appropriate methods for low level waste management investigations. Frequently the strata will be comprised of subareas in an overall region of interest. Although the usual textbook model for the method considers that an overall estimate (usually a mean or a total) for the entire area is the primary goal, it may often be true that separate estimates for the individual strata are also needed.

Individual strata may be considered separately simply because there is a need to have an independent estimate, or there may be reasons to compare

the several strata, e.g., to determine which have higher concentrations of radioactivity. In either case, it will probably be desirable to select the strata so that each is a self-contained unit, i.e., the elements are contiguous. It should be emphasized that this is not necessary if only totals or means for the entire area are wanted. In that case, a stratum may contain elements that are scattered about in the total population of interest.

The question of allocation of samples to strata is one that needs careful attention in survey design. By this we mean that a total overall sample of size n has to be broken down into subsamples assigned to each of the strata. If the main interest is in estimating a total or mean across the entire study area, one approach may be taken, while if an important aspect of the study is comparisons, yet another allocation is needed. These allocations can be very different. If the sampling is to be cost-effective, then it may be necessary to include a cost function in the design. Usually such a function is a linear combination of cost per element and sample size, but there is also a need to consider such additional features as travel time and/or sampling effort.

The basic methods of allocation, as given in textbooks, are proportional allocation, in which samples are distributed to the individual strata simply in proportion to the size of the stratum, and "optimum" allocation, in which samples are allocated in proportion to the product of size and the variance (sometimes standard deviation) of the variable of interest in the stratum. Very often, optimum allocation is by far the most efficient approach. It should be noted that proportional allocation is approximately equivalent to the allocation that would be obtained if the sample were simply distributed at random over the entire area. In general, the three approaches can be ranked by an inequality which indicates that the smallest variance (or narrowest confidence limits) is achieved by optimum allocation, with proportional next, and simple random sampling yielding the largest variance. It is important to note that such a result holds only if the design is correct, i.e., major errors in stratification can yield higher variances than would be obtained by simple random sampling. "Optimum" allocation is truly optimal, in a mathematical sense, only if the true stratum variances are known. In practice, one must estimate these variances before sampling takes place.

Cochran (1977:98) suggests three general rules for stratified sampling. These say essentially that the larger strata get more samples, that more variable strata require more samples, and that one should use relatively larger samples in a given stratum if sampling is cheaper there. Since radionuclides frequently follow a frequency distribution that is approximately lognormal, it is essential to take that fact into account in allocating samples for assessment of concentrations of radioactivity. Thus Eberhardt and Gilbert (1976) proposed three distinct allocations might be used in the appropriate circumstances:

1. Allocation proportional to areas of the strata.

2. Allocation proportional to fractions of the total quantity of radioactivity falling in each stratum.
3. Allocation proportional to product of area and square of the mean concentration in a stratum.

The matter of allocation and design of stratified sampling is sufficiently important to low-level waste management that a great deal more evaluation is needed than can be furnished here. Additional research on the topic will thus be needed to supply the necessary details for application to low-level waste site management.

4.2.5 Double-Sampling

As noted in Section 4.2.1, double-sampling involves the efficient utilization of data collected on two well-correlated variables, one inexpensive and readily collected for a large sample, while the other is appreciably more accurate, but relatively expensive to obtain. When the inexpensive measure can be collected on every sampling unit in the population, ratio or regression methods can be used. Usually, one cannot obtain data on every unit in the population so that an approximate approach has to be used, and this is double-sampling or two-phase sampling. Readers should note that there is at least one other method known as double-sampling. In quality control work, one may elect to take an initial sample and examine it for fraction defectives. If there are no defectives, or some small number in the sample, then the entire lot of product may be passed (accepted). If there are a larger number of defectives, then the lot may be rejected, or every element examined. However, if there is an intermediate number of defectives, a second sample may be needed for a decision. Methods for deciding on the several limits and other details of this process are known in "acceptance" sampling as double-sampling. There is no important connection with the approach being considered here, and known in survey sampling theory as double-sampling, beyond the obvious fact that both methods use two samples.

Double-sampling can also be used in an approach to stratified sampling. In this circumstance, an auxiliary variable may not be measured. A large initial sample is somehow classified into several strata, and a subsample of each stratum is used to measure the variable of prime concern. We then have a stratified sample in which the total number of units in each stratum is known only approximately, not exactly, as in the usual application of stratified sampling. In some circumstances, it may be possible to decide which stratum a unit should fall in without a measurement. In others, one may take a measurement. An example would be the use of a field survey radiation detector to take a quick measurement. Presumably the measurement could be used to erect strata or to carry out the ratio approach.

Double-sampling is mainly known as a means to improve efficiency of sampling when an estimate of a total or a mean is the main goal of the sampling. It can, however, be effectively used in an analytical approach, i.e., measurements on an auxiliary variable may substantially

improve the precision of comparisons of some sort. We have used both approaches in different applications. Gilbert and Eberhardt (1976) used double-sampling to estimate mean plutonium concentrations in surface soil, while Eberhardt (1978) gives some details of an analytical application.

Since double-sampling involves an approximation, some rules concerning circumstances of its application are needed. One rule of thumb suggests that a sample size of 30 units is needed. In some applications, a sample this large may be difficult to achieve, so there is a need to look into the consequences of using smaller samples. We have initiated a simulation study of the impacts of smaller samples aimed at determining when smaller samples might be utilized.

4.2.6 Repeated Sampling in Time

Most of the discussion of sampling in this preliminary report is couched in terms of a single sampling occasion. In reality, there will be various needs for repeated sampling of the same population. Under some circumstances, it may be desirable to keep the same sampling units throughout all of the study, while in others it may be desirable to draw a completely new sample each time. In most cases, some compromise between the two extremes may be the best strategy. With this wide a range of possibilities, the details of a given approach will have to be tailored to circumstances. Hence, in the present report, we only list the general rules, as given by Cochran (1977:344 ff.): Given the data from a series of samples, three kinds of estimates may be desired:

1. The change in a mean from one occasion to the next;
2. The average value over all sampling occasions; and
3. The mean for the most recent occasion.

Depending on which kind of estimate is needed, one has a choice of sampling strategy for the next sample taken from the population:

1. For estimating change, it is best to retain the same sample throughout, i.e., we assess the same sampling units on both occasions.
2. For estimating the average over all occasions, it is best to draw a new sample of units on each occasion.
3. For current estimates, either process gives about the same precision, and more detailed considerations usually lead to an intermediate strategy.

4.2.7 Sample Size

Methods for determining sample size depend very much on the sampling approach, and thus need to be considered in more detail in conjunction with the several different kinds of sampling and with the particular

subset of techniques being used. Consequently, we will not attempt to review the various prescriptions here. It is worth noting, however, that the broad applicability of the lognormal distribution as a model for radionuclide concentrations suggests that coefficients of variation will be constant over a wide range of concentrations. This means that an estimate of the coefficient of variation can be used to prescribe sample sizes for a given relative precision without advance knowledge of concentrations to be encountered in a given situation (cf. Eberhardt 1976).

4.2.8 References on Descriptive Sampling

As noted above, a wide range of books dealing with sample surveys are now available, so that the standard techniques are available for ready reference. Here, we list some of these references and give a brief indication of their nature. Specific references to these texts and to the extensive other literature on sampling are given throughout this report. Most of the available books and monographs are concerned with a relatively narrow range of survey sampling and are not as relevant to the problems encountered in low-level waste management as might be desirable. Much useful information is scattered about in the literature of entomology, fisheries, forestry, animal and plant ecology, and wildlife management and other fields. Locating and assessing relevant material in this rather diffuse literature will take an appreciable effort and is thus necessarily largely deferred to the next segment of the research in the present program.

Bailar, B. A., and C. M. Lanphier. 1978. Development of survey methods to assess survey practices. x+117 pp. American Statistical Association, Washington, D.C.

This is largely concerned with a developmental and feasibility study to assess the bases for present day survey sampling and is mainly concerned with interview surveys. The study has some relevance to low-level waste management in that some work of the kind considered in this book is needed in assessing the demographic aspects of a prospective waste repository site. However, the scope of the study considered in this book is substantially beyond what needs to be considered for low-level waste management.

Cassel, C., C. Sarndal, and J. H. Wretman. 1977. Foundations of inference in survey sampling. vii + 192 pp. J. Wiley and Sons, New York.

This is a somewhat technical book concerned with the process of inference from sample to population. It is of particular interest in the context of the present study in discussing the distinction between the finite population approach and the superpopulation approach. In the simpler, finite population approach, there is a single fixed population of units, each of which can be assigned a label (a number or numbers denoting the location of that unit in the population as, for example, a

population of canisters of waste). A sample can be drawn by various means, and a measurement made on each unit. It may be desirable to go beyond this and define "superpopulations" in various circumstances. One is when the measurement on a unit cannot be made without error, as when a subsample is assayed, or when a radiation detector is used. A new sample using exactly the same units would then give a somewhat different result. Hence there is some kind of superpopulation about which inferences may be made (this situation is not, however, studied in the book referred to here).

Another situation is that in which several samples are taken at two or more points in time. We might, for example, sample for radioactivity at a repository at intervals of a year. Usually it turns out that the measurements on sampling units are fairly well correlated in time, i.e., a high unit on the first occasion can be expected to be relatively high in the next sampling, and a low unit will remain relatively low. If this is true, then a rather smaller sample may be taken on the second occasion by taking advantage of the correlation in time. Assessing such data is facilitated by a superpopulation model, i.e., a very large population of possible outcomes on the second occasion can realistically be postulated, depending on the process that generates a change between the first and second outcomes, such as fallout from weapons testing.

Yet a third version of a superpopulation is described in Section 4.5 (Sampling for Pattern), in which a stochastic (probabilistic) process produces local "fluctuations" that are superimposed on an overall trend. Only a single instance is observed in any given situation, but analysis is facilitated by considering that an infinite number of realizations might occur. This approach may appear somewhat artificial, depending on one's point of view. Another possibility for generating a superpopulation is to produce a model that summarizes prior knowledge or subjective beliefs about a population in some fashion, leading to the "Bayesian" approach to statistics.

Cochran, W. G. 1977. Sampling techniques. xvi + 428 pp. J. Wiley and Sons, New York. Third Edition.

Cochran's book on sampling is a major reference in the field and first appeared in 1953. A second edition appeared in 1963, and is very widely cited. In our opinion, it is the best general reference for the needs discussed in this report. Since we refer repeatedly to this text, it will not be discussed further here.

Deming, W. E. 1943. Statistical adjustment of data. J. Wiley and Sons, New York (available in a Dover reprint, 1964).
Deming, W. E. 1950. Some theory of sampling. J. Wiley and Sons, New York (available in a Dover reprint, 1966).
Deming, W. E. 1960. Sample design in business research. J. Wiley and Sons, New York.

Deming's works provided one of the earliest comprehensive accounts of sampling methodology and theory, and his more recent book is widely used in various commercial activities. He has had a profound influence on the acceptance of sampling methodology in courts of law, and his contributions in this area may be of some importance in the low-level waste management context. Deming (1950:Ch.7) gives a useful distinction between enumerative and analytical uses of survey methods.

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

This is one of the few books in the environmental field that refers directly to sampling designs. It is, however, principally concerned with the assessment of environmental impacts. The focus is thus more on the lines of analytical sampling than survey sampling. The author recommends setting up a hypothesis and attempting to collect data to test that hypothesis against some specific alternative using a specific test of significance. He thus prescribes the use of controls and replications, but unfortunately the presumed replications in most instances are instead subsamples, making the tests of significance of questionable value. As noted in Section 4.3.4, we believe that research to delineate suitable alternatives should be an important part of the work done in the balance of the present study.

Hajek, J. 1981. Sampling from a finite population. v + 247 pp. Marcel Dekker, Inc., New York.

This monograph by Professor Hajek provides a theoretical basis for finite sampling, and should be a useful reference for technical investigations in that area.

Hansen, M. H., W. N. Hurwitz, and W. G. Madow. 1953. Sample survey methods and theory. Volume I. Methods and applications. xxii + 638 pp. J. Wiley and Sons, New York.

One of the basic references for sample survey methods has been this two volume work (Volume II provides the theoretical basis for Volume I). The first three chapters give a valuable non-technical review of the standard methodology for surveys of human populations and various kinds of establishments. Two of the authors were associated with the Bureau of the Census for many years so that many of the examples deal with the kinds of sampling programs relevant to that agency.

Jessen, R. J. 1978. Statistical survey techniques. vi + 520 pp. J. Wiley and Sons, New York.

This is largely a textbook, covering the main features of survey sampling, with emphasis on both surveys of people and businesses and in agriculture.

Kish, L. 1965. Survey sampling. ix + 643 pp. J. Wiley and Sons, New York.

The emphasis in this book is also on samples of human populations and social studies, supported by the author's long association with the Survey Research Center at the University of Michigan. There is a great deal of material on the utilization of complicated sampling "frames", necessary in the kinds of studies dealt with at the Survey Research Center, and on the practical aspects of designing and carrying out a large scale survey.

Mendenhall, W., L. Ott, and R. L. Scheaffer. 1971. Elementary survey sampling. vi + 247 pp. Wadsworth Publishing Co., Belmont, California.

An elementary description of the major sampling methods is given, along with the basic estimation and variance equations.

Raj, D. 1972. The design of sample surveys. xii + 390 pp. McGraw Hill, New York.

This is again a book mainly concerned with human populations, but with the broad basis engendered by association with the United Nations. A compact discussion of sampling methods is supplemented by extensive material on planning and executing surveys and an array of examples ranging from agriculture through demography, health, industry and commerce.

Slonim, M. J. 1960. Sampling. xiii + 144 pp. Simon and Schuster, New York.

This little book was originally published under the title, "Sampling in a nutshell." It does an excellent job of giving a thoroughly non-technical sketch of the major sampling methods.

Stuart, A. 1976. Basic ideas of scientific sampling. 106 pp. C. Griffin and Co., London. Second edition.

This is another elementary book, but it seeks to explicate the principles of modern survey sampling methods through simple examples and discussion of the underlying principles. An effort is made to adhere as closely as possible to the formal principles underlying sampling methodology.

Sudman, S. 1976. Applied sampling. x + 249 pp. Academic Press, New York.

An elementary and basic book on principles and methods for survey sampling, mainly aimed at the social sciences, and small-scale sampling.

Sukhatme, P. V., and B. V. Sukhatme. 1970. Sampling theory of surveys with applications. xvi + 452 pp. Iowa State University Press, Ames, Iowa. Second Edition.

A textbook and reference for sample surveys, with particular emphasis on agriculture. Both of the authors have been concerned with agricultural surveys, particularly through the Food and Agricultural Organization of the United Nations, and in India and the U.S. Due to the emphasis on agriculture, this book may be more useful in the context of the present study than will some of those references concerned mainly with surveys of human populations. The mathematical exposition is also useful as an alternative source of proofs to, for example, Cochran's book.

Williams, W. H. 1978. A sampler on sampling. xv + 254 pp. J. Wiley and Sons, New York.

Another elementary book, giving very lucid explanations of basic sampling, illustrated by numerical examples.

Yates, F. 1981. Sampling methods for censuses and surveys. xvi + 458 pp. MacMillan Publ. Co., New York. Fourth Edition.

This has been regarded as one of the major references on sampling, having been so regarded since the first edition in 1949. It provides extensive coverage of methodology and problems with a minimum of mathematics, and covers subject matter fields from agriculture to a worldwide survey of human fertility. If a two-volume library on sampling is to be selected, then it would most likely need to contain Cochran's book and this one.

4.3 ANALYTICAL SAMPLING

4.3.1 Introduction

Data collected by sampling can be used in an effort to infer something about the processes that led to the particular configuration of the population being sampled, or as a basis for comparing different subsets of the population. We emphasize the fact that such analyses can be quite risky when carried out on data collected by sampling (as contrasted to data collected in a properly designed and randomized experiment). These kinds of concerns led us to describe one kind of arrangement useful in environmental impact studies as a "pseudodesign" (Eberhardt 1976). In quite different circumstances (research in education and social sciences), a similar approach has been labelled as "quasi-experimental" (Campbell and Stanley 1966), which is probably a better term.

It is important to emphasize that the analytical methods and mathematics may not be much, if any, different whether or not randomization is feasible. A useful description is that of Scheffe (1959:3):

"The analysis of variance is a statistical technique for analyzing measurements depending on several kinds of effects operating simultaneously, to decide which kinds of effects are important and to estimate the effects. The measurements or observations may be in an experimental science like genetics or a nonexperimental one like astronomy. A theory of analyzing experiments naturally has implications about how the experiment should be planned or the observations should be taken, i.e., experimental design. Historically, the present technique of analysis of variance has been developed mainly in connection with problems of agricultural experimentation."

It should be noted that Scheffe's definition of one of the main analytical techniques now used makes no distinction between experiments or surveys in accord with the fact that there is often nothing in the underlying mathematical model or in the subsequent analysis to distinguish between the two possible sources. The difference is one of inferences based on the outcome of the analysis. However, it should be noted that some care has to be used in deciding what underlying model is appropriate, and hence what structure will be assumed in the formal statistical analysis. An example is provided by considering a situation in which it is necessary to test for movements of radioactivity within the boundaries of a waste site. In most cases, there will be a prevailing wind direction, so that the investigator might choose to lay out study plots in the upwind and downwind directions from the waste trenches, and sample these for contamination (various media might be considered). The actual sampling will have to be done as subsampling, i.e., material will be collected from trees, soil cores, etc. within the subplots, which may be randomly located as suggested in Figure 4.2. It is quite common in the analysis of such data to regard these subplots as replicates and to carry out an analysis of variance on this basis. This is, however, not a legitimate approach, and an error term based on subsampling will lead to unrealistic inferences. The situation can be complicated further by supposing that repeated samples may be taken from the subplots, so that time may appear in the analysis as an additional factor. If the collections in the subplots are randomized for each time of collection, a somewhat more realistic model for an analysis of variance may be proposed (as contrasted to the situation where the same locations within subplots are used repeatedly). In this case it may be realistic to regard the time periods as providing some sort of replication of the processes resulting in redistribution of radioactivity. A fuller assessment of such situations is not feasible here, but is planned as a part of the continuing research on this project.

4.3.2 Multiple Comparisons

In cases where field experiments can actually be conducted at low-level waste sites and where several treatments (i.e., different kinds of trench covers) are randomly applied, the methods of statistical analysis

FIGURE 4.2. Subsampling plots at a hypothetical low-level waste shallow land burial site. Differences in mean radionuclide concentrations on the two plots can be statistically evaluated, but cause cannot be safely inferred. When additional samples are obtained at other times, we recommend a new random selection of subsamples at each plot.

currently available (principally from agricultural experimentation) may generally be appropriate. For instance, the results from a designed and replicated trench cover study could be evaluated via analysis of variance. As noted in the introduction, inferences are limited to the site under study and the time of evaluation so that further experiments at a sample of other sites and repeated observations over years may be advisable.

In the monitoring situation depicted in Figure 4.2, subsamples from several plots, stations or streams may be obtained and periodically the question of differences in radioactive or chemical content may arise. If subsamples from each plot are separately analyzed (i.e., not composited, Section 3.2) and the resulting data set does not contain less than detectable concentrations (Section 3.6), then standard statistical methods may again be appropriate. However, an inferential problem

may arise when statistically significant differences are detected (i.e., what caused the elevated activity or where did it come from?).

In both the trench cover experiment and question of differences among monitoring stations over a site, a statistical comparison of all treatment or station means may be needed. When some of these are planned in advance, either the least significant difference or orthogonal contrasts in the analysis of variance can be used (Steel and Torrie 1980). We envision few circumstances at low-level waste sites where comparisons will be planned in advance. However, statistical methods are available for "testing effects suggested by the experimental outcome." These are generally known as multiple comparisons.

Of those in current use, Scheffe's method (Scheffe 1959:66) is the most conservative (i.e., larger differences between two means will be needed to declare a significant difference at a given probability level). Its conservatism may be an advantage for low-level waste site monitoring studies, since Steel and Torrie (1980) advise "the test's use seems most appropriate for 'data dredging' -- looking at contrasts suggested by the data -- as is often done in survey analyses." Another advantage is that the theory has been developed for circumstances when unequal replications (subsamples) per treatment or station are encountered. For two other common methods [Student-Newman-Keuls, (Newman, 1939), and Duncan 1955] the validity of the approximations suggested for use with unequal replication has not been verified. Recent work (Dunnett 1980a) suggests that the approximation for Tukey's test (Kramer 1956; see Tukey 1953) is valid.

In the case of designed field experiments, the conservative nature of Scheffe's test may result in low statistical power (Steel and Torrie 1980 :183). Thus, it appears that Tukey's test could be recommended as the next "most conservative." Its principal departure from Scheffe's test lies in the fact that only "pairwise" comparisons can be made. In contrast, Scheffe's method allows testing all possible contrasts and the construction of confidence intervals for the corresponding linear functions of parameters.

For illustration, we have calculated the critical value or values needed to declare pairwise comparisons of a set of six means (replicated five times each) significantly different ($P < 0.05$) after a one-way analysis of variance (Table 4.1). The different methods result in different critical values because of the intended scope of each test, whether or not experiment-wise, per comparison-wise, or family of hypothesis-wise error rates are assumed and other factors.

In a recent paper, Stoline (1981) recommends the use of Tukey's (1953) test for simultaneous estimation of pairwise differences in balanced one-way design and the Kramer (1956) modification for unbalanced cases (assuming homogeneity of population variance). For balanced data sets where heterogeneity of variance exists, he advises the procedures suggested in Dunnett (1980b). For unbalanced cases where heterogeneity

TABLE 4.1. A Comparison of the Different Results Obtained Using Various Multiple Comparison Tests. Values in the Difference Column are the Differences Among Pairs of Means Needed to Declare Them Significantly Different ($P < 0.05$)⁽¹⁾

Test	Difference
Least Significance Difference	4.5
Scheffe's	7.9
Tukey's	6.7
Student-Newman-Keuls ⁽²⁾	4.5-6.7
Duncan New Multiple Range ⁽²⁾	4.5-5.1
Waller-Duncan's Bayesian k-ratio t-test ⁽³⁾	4.2

1. Experiment consisted of six treatments of 5 replicates each. We assumed a fixed effects model, normally distributed data and used a one-way analysis of variance.
2. These tests depend on the magnitude of mean differences when they are in an ordered series.
3. Based on Type I to Type II error seriousness ratio (Dixon and Duncan, 1975).

exists (the likely circumstance at low-level waste sites), little is known about optimum tests. If appropriate transformations do not "adjust" the data then it appears that additional research may be needed.

Since the statistical analysis of low-level waste site data could encompass many of the complications above, it appears that the graphical procedures for multiple comparisons suggested by Hochberg et al. (1982) may be very useful for judging differences among monitoring data sets especially by site operators. We plan continuing research on this methodology. If we find it is appropriate, we will prepare a computer program to implement the technique.

4.3.3 Statistical Evaluation of Data Collected for Analytic Purposes

As pointed out in the introduction to this Section (4.3.1), the assignment of cause/source generally cannot be made from the finding of "statistically significant differences" in radioactivity at locations over a low-level waste site, when a design such as shown in Figure 4.2 is used. One possible approach is discussed in the next Section (4.3.4).

Examination of radionuclide monitoring data collected at the Maxey Flats, Sheffield, and Beatty sites indicated that the air, ground-water, surface water, soil and biotic samples (when collected) are either collected on a routine (biweekly, quarterly) or sporadic schedule at

stations that are located in areas just near the site boundary or are not far off-site. Sometimes data are not collected because of equipment malfunctions. Routine reports containing compilations are published and little is said about the prior period or how current and prior results interrelate. Generally no subsite or subplot samples are taken (i.e., one sample/location).

Tritium concentration as a function of time in Maxey Flats water is plotted in Figure 4.3 (obtained from Appendix 7 of Dames and Moore (1982); which in turn is Appendix 6 of The Commonwealth of Kentucky (1982)). It appears that these monthly data, in concert with concurrent stream flow estimates, will eventually be used to estimate tritium export from the site. We note that these data were accompanied by analytic error estimates (two standard deviations were about $\pm 5\%$ of the plotted values). However, analytic error estimates cannot be used in statistical tests where the objective is to assess differences (due to location) among stations (more than one sample at each station is needed). We note too, that the data set is almost complete in that there is only one missing value (May 1981, Station A). It is impossible to ascertain the variance of any single point estimate since only one sample was taken at each station/time. However, visual appraisal indicates that tritium concentration at Station B was usually highest while Station A was always lowest and Stations C and D intermediate. Even though these observations are rather convincingly supported by the data, statistical evidence would be difficult to obtain. Peaks and valleys may be associated with some other parameters measured (i.e., rainfall, surface run-off, etc.) or location relative to an evaporator being run concurrently on-site. If the question is "are there statistically significant differences among stations?" (cause not of concern), then Thomas et al. (1978) offer several prospects for qualitatively evaluating the data. Even though these methods of quantitative/qualitative analyses have severe limitations, they can be used to suggest additional monitoring. However, even when results from more than one sample at each station-time are available, significant differences still may simply be a function of where and how samples were obtained, differential binding to sediments or soils prior to arrival at the sampling stations (probably not for tritium), or a myriad of other factors. In other words, differences in tritium concentration at stations could be related to a myriad of explanations.

Data for some of the other radionuclides measured are in Table 4.2. These surface stream samples are collected quarterly (Appendix 5 of Commonwealth of Kentucky 1982 and a personal communication from Mr. Doyle Mills of the Commonwealth). Results from only one collection date (8/4/82) are presented in Table 4.2. We note that the less than values reported are actually the lower limit of detection which can cause special problems during statistical analyses (Section 3.6). Again, only one sample was obtained. Since the rationale for collecting the data is not given, we suppose that for any radioisotope the questions might be: 1) are the concentration values obtained at each station significantly different, 2) are differences consistently associated with location, or 3) does the concentration observed at any location exceed a prescribed limit? The data in Table 4.2 cannot be used to answer (in a statistical

FIGURE 4.3. Tritium concentration in water samples obtained at four Maxey Flats stations as a function of time. Only analytic (counting) error estimates were available ($\pm 5\%$, not shown) so no estimate of location variance could be obtained. Thus, the apparent differences exhibited at station A and B cannot be statistically evaluated without ambiguity. More than one sample should be collected at each location (at one or more times of interest) to allow statistical evaluation. Our experience indicates that percentage error (coefficient of variation) for replicates at locations can range from 15 to over 100% (depending on the type of sample and radionuclide) compared to analytic errors of 1 to 10%.

sense) those questions because only one observation was obtained per station/time. We note that consistent differences in radionuclide concentration between two stations over time may be evidence for a real difference but, there is no statistically unambiguous way to test the difference hypothesis. If subsamples had been obtained, then a statistical analysis of all the quarterly data sets (replicated, but perhaps with missing observations) could be approached via modification of the methods for "messy data" outlined in Urquhart and Weeks (1978). Urquhart's suggestions would be particularly attractive if a concurrent analyses of related radionuclides was needed (i.e., all transuranics).

TABLE 4.2. Radionuclide Concentrations in Stream Sediment Samples Collected at Maxey Flats, Kentucky on August 4, 1982. This table represents a common form of data presentation. It should be noted that there is no indication that more than one sample was obtained at any sampling station or that these are results from composite samples. It is, therefore, unclear as to just what question the data collection scheme was designed to answer, and how they might be used in a statistical analysis (since variance estimates cannot be made). Also noteworthy are the numerous less than (<) estimates which represent the lower limit of detection. These would complicate a statistical analysis even if more than one sample per location had been collected (Section 3.6).

Station	pCi/gm dry $\pm$ Counting Error										
	^{90}Sr	^{89}Sr	^{60}Co	^{137}Cs	^{40}K	^{226}Ra	^{235}U	^7Be	^{95}Nb	^{238}Pu	$^{239}\text{Pu}/^{240}\text{Pu}$
SHD4A	0.08 $\pm$ 0.06	<0.11	<0.024	0.17 $\pm$ 0.02	3.7 $\pm$ 0.37	0.63 $\pm$ 0.06	<0.13	<0.39	<0.051	0.82 $\pm$ 0.36	<0.14
SHD6	0.24 $\pm$ 0.08	<0.16	<0.11	0.27 $\pm$ 0.07	17.0 $\pm$ 1.7	2.3 $\pm$ 0.23	<0.54	<0.83	<0.12	0.45 $\pm$ 0.39	<0.17
SHD13A	0.98 $\pm$ 0.17	<0.37	0.57 $\pm$ 0.09	0.58 $\pm$ 0.08	21.0 $\pm$ 2.1	3.0 $\pm$ 0.30	<0.63	<0.79	<0.12	<0.33	<0.22
SHD033A	<0.12	0.24 $\pm$ 0.17	<0.11	0.19 $\pm$ 0.07	20.0 $\pm$ 2.0	2.7 $\pm$ 0.27	0.71 $\pm$ 0.43	1.3 $\pm$ 0.68	<0.20	0.35 $\pm$ 0.22	<0.89
CK-AQS-Sediment Main Drain	1.57 $\pm$ 0.16	<0.35	0.42 $\pm$ 0.07	0.52 $\pm$ 0.07	15.0 $\pm$ 1.5	0.96 $\pm$ 0.11	NA ²	<0.71	<0.08	0.83 $\pm$ 0.031	<0.21

¹Sampled 7/26/82, all others 8/4/82

²NA = Not Analyzed For

The assumptions implicit in the use of many "classical" statistical methods usually include additivity of terms in a linear model, homogeneity of variances, normality of data and independence of observations. In all the low-level waste site monitoring programs we have assessed, little consideration was given to any of these factors when deciding how to collect or evaluate data. In our experience with biological monitoring programs, the most frustrating problem is the failure to consider the lack of statistical independence of observations taken as a function of time. The lack of independence is evident in Figure 4.3 since it is likely that successive observations at each station are correlated. There were no low-level waste site data sets where enough apparent cycles were available to use time series methods (Box and Jenkins 1970) so that statistical independence would not be a problem (at least 8-12 cycles are needed) or where linear or non-linear regression might be appropriate. Eberhardt (1978) suggests that additivity may be the most important statistical factor to consider and suggests a logarithmic data transformation to attain this goal because the error structure of the underlying probability model appears likely to be multiplicative. Moreover, he also shows that achieving normality may be another desirable result of such a transform, but that should not be the central reason for transforming data. The validity of some of the methods of analysis suggested in Thomas et al. (1978) can be questioned as a consequence of Eberhardt's (1978) work.

We presume that future monitoring data collected at new low-level waste sites will, at the very least, be collected as depicted in Figure 4.2 (i.e., subsamples obtained at each location and perhaps re-randomized as suggested in Section 4.3.1) and that some of the ideas expressed in the next section can be incorporated, if causal factors are to be ascertained. Our plan is to provide guidance in the next phase of the project. Some possible examples for analytical surveys are in Sedransk (1965a, 1965b, 1966).

4.3.4 The Question of Cause

It appears to us that the need to address questions about causes of radionuclide redistribution and/or migration at low-level waste sites is not mandated in 10 CFR 61.53(c) (Section 1.2). However, monitoring for detection of such events is required. As pointed out in Section 4.3.1, simply using subsamples of soils at various locations near the site fence or wells off-site can detect below and aboveground off-site transport simply by comparison to subsampled "control" areas within the site. However, this approach does not necessarily define source and/or cause. Since likely aboveground vectors are wind or surface water, a scheme such as that depicted in Figure 4.2 (where new replicate soil samples are chosen on each sampling occasion) might identify atmospheric forces. Similar samples at usual surface runoff sites could also be compared to some "control" in an analogous fashion. Rather subtle (see 10 CFR 61.53 c and d) radionuclide movement might be detected if data prior to an event (i.e., upwind/downwind ratios) are compared to results obtained after an event. The event might just be the occurrence of a high concentration somewhere on-site or a catastrophic wind or rainstorm. While the idea is not new (Eberhardt 1976 called the procedure a

pseudodesign), and may not be appropriate for all new low-level waste sites, some exploratory research may be warranted to assess its statistical power for sites in high rain/wind locations.

We also note that when a routine collection results in an unacceptably high measurement for soil, water, biota or air, additional studies are generally conducted. The foregoing scenario almost always results when unacceptably high readings are obtained at any installation, nuclear or otherwise. The questions of how to design monitoring programs that implicate a cause, or how to monitor for source (after a high value is obtained), are not specifically addressed in 10 CFR 61. Moreover, 10 CFR 61 states that after an "early warning," mitigative measures may be instituted. Specific sampling needed to assess the source location is only implied in the requirement that "mitigative measures must be instituted."

4.3.5 References on Analytical Sampling

A wide range of books dealing with sample surveys are listed and described in the preceding section on descriptive (survey) sampling. Unfortunately, analytical sampling has not received much attention in the literature, possibly because it falls between two fields, that of statistical analysis and that of survey sampling. Some of the survey sampling texts do, however, have short sections on analytical sampling, and these will be briefly discussed here.

Cochran (1977:4) distinguishes the two methods:

"Sample surveys can be classified broadly into two types -- descriptive and analytical. In a descriptive survey the objective is simply to obtain certain information about large groups: for example, the numbers of men, women, and children who view a television program. In an analytical survey, comparisons are made between different subgroups of the population, in order to discover whether differences exist among them and to form or to verify hypotheses about the reasons for those differences."

Cochran provides several brief comments on features of sample surveys that concern analytical purposes. The most extensive is a short section on strata as domains of study, i.e., surveys designed to make comparisons among strata selected for analytical purposes. He supplies formulae for allocating samples to strata in this case.

In his pioneering book on sampling, Deming (1950) makes essentially the same distinction between descriptive and analytical sampling, and gives a sizable number of examples. He also makes the interesting point that, for analytical purposes, even a complete tally of all of the units in particular domains of study will not negate the use of statistical methods for analysis, whereas in a descriptive study, making a complete tally eliminates the need for any further analysis. The reason, of course, is that the analytical survey is conducted with the notion of trying to infer from observations why the population takes on its

particular characteristics. Consequently, the superpopulation idea dominates for analytical purposes, i.e., we suppose that the particular population under study arises from an infinite (or nearly so) array of populations (the superpopulation) that might have resulted from the processes creating it. This is the reason that the finite population correction, used in descriptive sampling, is dropped for analytical purposes.

Jessen (1978) briefly discusses analytical surveys. He suggests a classification for design purposes that is essentially that used for experimental designs, i.e., by factors and levels. One may thus attempt to classify the population being sampled by certain factors that can be subdivided into levels, and then subsample within these categories. One might thus wish to contrast some feature of low-level waste sites with reference to the several classes of waste (as defined in 10 CFR 61) and seek to sample each of several levels (in terms of quantities of waste "ticketed" into the site). As mentioned previously here, the problem is that the investigator can only use sites that are available, and has none of the control available in an experimental set-up. It is thus advisable to keep any attempt to set up such a framework as simple as possible.

Kish (1965:Ch.14) provides a discussion of issues of inference in survey sampling. He suggests that four classes of variables may be present in a given situation: (1) explanatory variables (also known as the experimental variables), (2) some extraneous variables that can be controlled, either in selection of units or by an auxiliary measurement, (3) other extraneous variables that are uncontrolled and confounded with the explanatory variables, and (4) the remaining extraneous, uncontrolled variables that can be treated as randomized errors. Often, the difference between an experiment and a survey depends largely on (3), since those variables can usually be combined into class (4) in the process of designing the experiment. He also notes that much of the statistical analysis of experiments is limited to inferences about the population encompassed by the experiment, which may be a much smaller population than that for which the investigator wishes to make inferences. There is thus an argument for somehow integrating the experimental and the sampling approaches into an overall course of investigation. We believe that this may be a valuable approach for consideration in low-level waste management, i.e., that experimental on-site studies ought to be designed so as to be included in a framework that encompasses extensive studies, by sampling, of the system of overall concern.

Yates (1981:Ch.9) provides the most detailed assessment of analytical surveys of any of the references examined thus far. He remarked, however, that "It must be emphasized, however, that the chapter is not intended to be an exhaustive treatise on the analysis of investigational surveys: this would require much more space than is available here." Yates has a good discussion of the distinctions between an experiment and a survey, and brings out the point made above with a useful example on agricultural experimentation. Paraphrasing it to relate to low-level waste management, we may suppose that, for example, various kinds of

treatments are to be tested experimentally to determine their efficiency in preventing wind (or water) erosion of trench caps.

Such an experiment may be conducted at a particular site. Given enough replication and a satisfactory design (randomized blocks, perhaps), quite definite inferences about effectiveness of a given set of treatments at that site are possible. Usually, these cannot be extended to all prospective sites, so further experimentation may have to be conducted at a sample of all likely sites. Also, there is abundant evidence, from agricultural studies over many decades, that experiments conducted in different years give different results. It thus becomes necessary to repeat the experiment in a number of years, again involving sampling, now in time.

One of the challenging problems needing further attention is thus how to construct studies that combine experiments with sampling in order to yield the most information at minimum cost. In the example suggested above, several of the prospective treatments will no doubt already have been applied at various times and places. Sample surveys might thus be conducted to make the fullest possible utilization of existing opportunities. So far as we know, this approach is virtually unexplored in a formal sense (but obviously the combination of experience and experiment is scarcely new, and has been discussed in the literature, for example, by Professor John Tukey).

Yates goes on to discuss a number of important topics in analytical sampling. He uses several examples from agriculture and population surveys, and addresses the use of ratios, simple linear regression, multiple regression, and the development and use of general linear models for survey data. Transformations of variables and the analysis of contingency tables are also discussed. The various kinds of biases in regression analysis and the pitfalls of interpretation are discussed in detail, using data from actual surveys. Yates' treatment is the most extensive of those available in survey sampling texts.

4.4 SAMPLING FOR MODELLING

4.4.1 Introduction

In many circumstances appropriate to low-level waste management, it will be desirable to sample in order to fit a model of some sort. In others, fitting a model may not be essential, but there is nonetheless a need to devise a sampling plan, and possible models for the process may be helpful in constructing such a plan. The kinds of activities considered appropriate in this section include monitoring biota, air sampling, and development of profiles of concentrations of radioactivity in soils with distance from a source, depth, or over time.

In most of the situations considered here, the main statistical problem is one of determining how to distribute samples over time (or distance, or depth). In practice, just one scheme has been utilized almost exclusively. That is simply to space the samples uniformly, e.g., to sample once a day, once a week, at one meter intervals, 5 cm depths,

etc. Probably the choice of uniform sampling is justified if there is no advance information on the phenomenon being studied. If there is advance knowledge of the likely pattern of events, then choosing evenly spaced samples may be a very bad choice, giving highly unsatisfactory results.

A simple example lies in sampling some process involving exponential decay. If the decay is quite rapid, most of the samples will be wasted, since the measured levels may be so low as to provide almost no information about the process. Actually, this is perhaps the only example where another sampling scheme has occasionally been used. A few perceptive individuals have chosen to sample uniformly on a logarithmic scale. This concentrates the sampling near the beginning of the decay, thus providing a concentration of measurements at times when levels are high. The two schemes (uniform and "geometric" spacing) are illustrated in Figure 4.4.

Over 20 years ago, the need to consider efficient sampling patterns for circumstances where a model is to be fitted was studied by Box and Lucas (1959). Their results were soon recognized as being useful in industrial experimentation, and a good deal of further work was done to develop and extend the early research. Unfortunately, the results obtained in the industrial area have largely gone un-noticed in other places where they could be useful. We have attempted to point out the possibilities and to develop methods for biological problems involving trace substances, including radionuclides (Eberhardt et al. 1976, Eberhardt 1978).

In the following sections, we review some of the methodology, available results, and further research needs relevant to low-level waste management. It should be remarked here that we are not attempting to directly consider large and complex simulation models here. About all that can be done in sampling for parameter estimation for such models is to proceed in a piecewise fashion, i.e., to isolate a component subsystem and deal with it separately. In reality, this is how virtually all such models are constructed in the first place.

4.4.2 The Box and Lucas Scheme

Box and Lucas (1959) considered a general model where a response variable is a known function of k variables ($x_1, x_2, \dots, x_k$) and of p parameters ($B_1, B_2, \dots, B_p$):

$$Y = f(x_1, x_2, \dots, x_k; B_1, B_2, \dots, B_p)$$

In its general form, the Box and Lucas scheme thus permits study of a multivariate system, with variables like time, temperature, concentration, etc., controlling the level of some product through a known functional relationship with fixed (but unknown) constants (parameters).

FIGURE 4.4. A hypothetical example of cesium-137 concentration as a function of soil depth which shows uniform versus geometric sampling for an exponential curve. The uniform sampling schedule (I) has only a few points in the depth interval of major importance and over samples the remaining interval, while the geometric schedule (II) concentrates most of the sampling in the important depth interval.

An example relevant to industrial applications is the experimental operation of a "pilot plant" in which a chemical process is being used to produce some product.

A comparable situation in low-level waste management may be an experiment in which various chemical states, chelating agents, oxidants, temperature, pressure, moisture, etc., are to be investigated with respect to behavior of wastes in a repository. In a sampling context, we are principally concerned with models for such things as concentrations of radionuclides with time, distance, or depth. An appropriate model will then be:

$$Y = f(t; B_1, B_2, \dots, B_p)$$

where t denotes time (but may alternatively represent depth or distance).

For a specific example, we consider a two-compartment model of radio-nuclide release from a source (waste depository perhaps) and retention in some receiving system, which might be a defined region of soil, or some element of the biota. We suppose that the input (y_1) decays away exponentially over time, so that the quantity in that compartment is (at time t):

$$y_1 = y_0 e^{-B_1 t}$$

We further suppose that the receiving system (y_2) loses a constant fraction of its contents per unit time (through decay and/or "washout" or excretion). The appropriate equation for quantity in the receiving system at time t is then:

$$y_2 = \left[B_1 y_0 / (B_1 - B_2) \right] \left(e^{-B_2 t} - e^{-B_1 t} \right)$$

We note that this equation is also essentially a "reaction-rate" model, so that the chemical analogy continues to be relevant. In the equation, y_0 denotes the initial quantity in the "donor" compartment which is transferred to the receiving compartment at rate B_1 . The receiving compartment loses the material at rate B_2 . In some circumstances, it is necessary to regard y_0 as a third parameter; in reaction kinetics, y_2 is usually expressed as a fraction of the (known) initial quantity of reactant (y_0) and thus only two parameters are shown in the equation and considered in estimation.

The objective of the study is to select the optimal times to observe (sample) the process in order to estimate the parameters for the equation. Normally there will be a defined time period in which the observations are to be made, either inherent in the conditions under which the study is conducted, or determined by the observer. These can be conveniently represented by $t = t_{\min}$ and $t = t_{\max}$, with $t_{\min}$ frequently (but not necessarily) the time at which the process is initiated.

The actual process for selecting optimal times calls for finding partial derivatives (f_{ru}) of the function, f (the model), with respect to the r^{th} parameter and the u^{th} sampling time. Suppose that there are to be N sampling times. We then represent the partial derivatives in an $N \times p$ matrix, F (N sampling times for p parameters). The optimal sampling times are those that maximize the determinant of the matrix product $F'F$ (i.e., the transpose of F multiplied times F), or, equivalently, minimize the determinant of the inverse of the same matrix product. To

obtain a numerical answer, it is of course necessary to specify parameter values.

On first encounters with the Box and Lucas scheme, an initial reaction is that it is not very helpful to require advance knowledge of the parameters that one is seeking to estimate. Two points must thus be emphasized. One is that the models to be considered are all non-linear in the parameters. Realistic study of such models requires some advance knowledge of the parameter values to be expected in given circumstances. Otherwise, the rapid changes associated with non-linearity are likely to leave the observer with virtually no useful results (since the action will either be practically over with or not yet started at the time the observer samples, unless there is some knowledge of approximately how the model behaves, and thus roughly the parameter values). The second saving feature is that most of the curves denoting optima are fairly flat-topped, so that an approximate advance knowledge of relevant parameters suffices to produce useful results.

Finding the optimal sampling times in an actual problem can be a rather difficult analytical exercise, and will require computerized searches in many instances. We will provide references to examples in the literature below. For a very simple case illustrating the mechanics of the method, consider a simple linear model:

$$Y = A + Bt$$

The partials with respect to A and B are unity and t, respectively, so the matrix of partials for N sampling times and the matrix product F'F are:

$$F = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_N \end{bmatrix}$$

$$F'F = \begin{bmatrix} N & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix}$$

A minimum for the determinant of $(F'F)^{-1}$ is obtained by maximizing $N\sum(t_i - \bar{t})^2$, and this is very simply achieved by splitting the sample

into two equal lots and concentrating the two halves at the beginning and end of the observation period, i.e., $N/2$ samples are taken at $t_{\min}$ and $N/2$ at $t_{\max}$. This is, of course, the result obtained by minimizing the variance-covariance matrix of the usual least-squares estimates of the parameters of the linear regression model.

Given normal, additive errors, the solution is exact for a linear model. Non-linear models necessarily involve the parameter values in the partial derivatives and thus only approximate solutions are possible. The optimum design is to select a set of sampling points that minimizes the determinant of $(F'F)^{-1}$, the inverse of the matrix product. This provides an approximate minimum joint confidence interval (confidence ellipsoid) for the set of parameter estimates.

4.4.3 Errors of Measurement

Box and Lucas (1959) consider only additive normal errors, whereas the distributions encountered in waste management studies are usually markedly skewed, and may usually be assumed to be approximately log-normal. It is very important to consider the effects of this difference in searching for optimal sampling times. For a simple illustration, consider a simple exponential model (the "decay-law"):

$$Y = Y_0 e^{-Bt} + e$$

where e now denotes errors of measurement, and is assumed an additive, normally distributed random variable, as is done by Box and Lucas. The optimum sampling times are then at $t_{\min}$ and at a time when about 37% of the initial value remains (i.e., about 63% of the substance has decayed away), according to Box and Lucas (1959:85).

If we now suppose that the errors are lognormally distributed and multiplicative, the usual procedure will be to transform the observations by taking logarithms. This gives the simple linear regression model used as an example above, in which we found that the optimum sampling times were $t_{\min}$ and $t_{\max}$. Clearly the two assumptions about error structure give different results. Most of the available literature considers only additive normal errors, so that it will be necessary to reconsider and revise the results and recommendations available there for use in low-level waste management studies. This is a component of the continuing work projected under the present program. We note, too, that this illustration is somewhat oversimplified, in that, in the second case (linearized model), $t_{\max}$ cannot be too large or the sampling errors will be excessive, due to decay of the substance being studied. One answer, of course, is to adjust counting times (when possible) so as to obtain comparable variances.

4.4.4 Additive and Multiplicative Parameters

One of the rather surprising outcomes of the Box and Lucas scheme is that the number of optimal sampling times (N) is equal to the number of

parameters to be estimated (p). This means that the $N \times p$ matrix of partial derivatives is a square matrix ($p \times p$). The search for optimal sampling times then depends on maximizing the absolute value of the determinant of this square matrix. This makes it fairly simple to show that the effects of "additive" parameters are such that it is not necessary to obtain an advance estimate of the values of such parameters. Hence, in a model such as:

$$Y = B_1 + f(t; B_2, B_3)$$

only the estimates of B_2 and B_3 are required in searching for optimum sampling times. We will not go into details here, but mention the point because it bears on the issue of multiplicative and additive errors discussed above. When a logarithmic transformation to achieve additive errors is used, the model is changed, usually resulting in some parameters becoming additive, and often simplifying the search for optimal sampling times.

4.4.5 Discriminating Between Models

Although the fact that the optimal set of sampling times equals the number of parameters is convenient in many respects, few practical-minded people will want to settle for such a limited number of sampling times. There are several reasons for such a view. Perhaps the most important is that we seldom can be absolutely sure that we actually know the correct model for the phenomenon under study. Another related point is that a wider set of points may be more useful in testing fit of the model and estimating variances. The main recommendation evolving from research on the topic over the last two decades has thus been to include some sampling points that have been determined as optimal for an alternative model, along with others chosen to give an efficient test for discriminating between alternative models. Further research concerned with models appropriate for waste management contexts will thus need to consider this feature along with those of the preceding two sections. In conjunction, they will provide directions and boundaries for that work.

4.4.6 Response Surface Models

A good deal of statistical work has been done on response surface models. Reviews by Hill and Hunter (1966), and Mead and Pike (1975) summarized some of this work and described response surface methodology. Various aspects of those studies may be of interest in waste management. It is important to note here, however, that response surface models are most often of interest and utility in searching for a maximum or a minimum. The focus is usually on finding an optimum set of treatment values to give maximum yield from some process. In the present context, a response surface model is used to find the set of times that locates a maximum of the curve (or some other feature, as a minimum). The Box and Lucas methodology, on the other hand, seeks to obtain optimal estimates of parameters in the model. The objectives are thus not the same. We

believe that response surface methodology may be useful in sampling for pattern, in common with trend surface methods.

4.4.7 References on Sampling for Modelling

The other sections providing references in this report have largely included only books on the subject matter under discussion. Unfortunately, there are few such sources on sampling for modelling. We thus include here a selection of published papers, and plan a more detailed review of the appropriate literature in the balance of the research program. Two textbooks (Beck and Arnold 1977, and Bard 1974) have fairly extensive discussions on design of experiments with nonlinear models. Chapter 8 of Beck and Arnold (1977) is particularly useful since it provides an accounting of a variety of criteria for optimal experiments. The focus is, however, chiefly on engineering.

In a biological context, the book edited by Endrenyi (1981) is especially important in a low-level waste context, due to the emphasis on kinetic models. Brief treatments of a range of relevant issues in modelling and model-fitting are provided, including a number of applications of the methodology described here. Another text, by Box, Hunter and Hunter (1978) contains somewhat less material on nonlinear models than those listed above, but is worth examining due to the fact that a philosophy and approach to experimentation and modelling is discussed in some detail. Since the authors of this book have been extensively involved over many years in developing and extending the Box and Lucas scheme, their general views are valuable in applications to a new field, as proposed in this report.

Some references to the journal literature relative to sampling nonlinear models may serve to provide readers with an impression of the material available. As noted above, a more detailed review of the literature is needed, and planned for inclusion in the next stages of the present project. A caveat and warning may be in order here. Virtually all of the literature that we have examined is directed to experimental studies, mostly in industrial research. The reader who expects to find references to sampling applications will thus be disappointed. As noted in the introduction to this section, we have begun to adapt these methods to field sampling, but much is yet to be done along these lines before there is a consistent and integrated methodology suitable for sampling needs.

As noted in Section 4.4.1, the pioneering paper in this area was that of Box and Lucas (1959). The next most important reference is likely to be the review by Cochran (1973). He traces the subject of experiments for nonlinear functions from some investigations of R. A. Fisher up to the 1970's, by which time much of the pioneering work had been accomplished, noting that research had largely been concentrated in two areas: (1) optimal estimation of the parameters, under an asymptotic criterion, and (2) tests of the adequacy of a model and discrimination between models, plus model-building. He reviews these areas briefly, thus providing a useful guide to work up to the time of his review.

One of the apparent limitations of the Box and Lucas scheme is the outcome that the number of optimal sampling times is equal to the number of parameters. This point was examined in some detail by Atkinson and Hunter (1968) who confirmed its validity and recommended that replications of the same set of points be used. They illustrate their recommendation with the model used by Box and Lucas (1959), and some others. M. J. Box (1970) also considered the replication aspect, and gave results of some further experience with the scheme. Box (1971a) examined an important practical problem, i.e., what to do when the experimenter is mainly interested in a subset of the parameters, and provided an approach. Hill and Hunter (1974), also consider this question and provide a design criterion for experiments concerned mainly with a subset of the parameters.

The limited scope of optimal sampling points provided by the Box and Lucas scheme can be extended by considering alternative models for the phenomenon under study, so that two or more sets of design points can be produced. Hill, Hunter, and Wichern (1968) approached this prospect by a sequential process, in which model discrimination is first emphasized, with a gradual shift to emphasis on estimating parameters for the "best" model as determined by initial experimentation. A good deal of attention has naturally gone into the question of discriminating among candidate models, due to its general importance in scientific and engineering experimentation. Hunter and Reiner (1965) proposed sequentially selecting experimental points at which two models differ the most. Box and Hill (1967) consider sets of conditions for the purpose. Atkinson (1969) focussed on significant differences in fits of the candidate models to the observations, and later Atkinson (1972) studied the use of extensions of a given model to test fit. These and many other references on model-building seem valuable as a source of ideas for developing ways to choose sets of points for extending the basic Box and Lucas scheme.

4.5 SAMPLING FOR PATTERN

4.5.1 Introduction

There are many ways to summarize a spatial pattern. Choosing among these possibilities depends on the purposes to be served and the assumptions that can be made about the underlying process that generated the particular pattern under study. It is important to have in mind the different aspects of a given situation. In the circumstances considered here, we may be required to deal with a single, specific area, and are given data on a particular set of samples previously taken over that area. The objective is to provide some kind of a summary of the pattern indicated by the samples, usually as a map. The issues that need to be considered depend on the uses to be made of the map.

In the frequent circumstances where the data are simply those that happen to be available, it is likely that there will be areas with very few points and others with clusters of points. Simply taking an average across all of the data may give a result that is quite seriously biased, i.e., differs substantially from the true (but unknown) mean value for

the entire region of interest. In this case, an alternative is to use the data to fit some kind of surface to represent the values associated with the available data points, i.e., to interpolate between data points. A mean value for this surface may then provide a better estimate of the true value. Also, such a map will provide the user with a much better "feel" for the nature of the dispersion of the variable of interest.

In the more satisfactory circumstances where sampling locations can be selected according to a design, it is possible and usually desirable to avoid clustering. An important decision may then be whether or not to resort to systematic sampling. As discussed in Section 4.2.2, this is an unresolved question, needing further research. Various alternatives involving restricted randomization require investigation, and serious consideration in designing a field study.

An intermediate set of circumstances also needs to be considered, in which a set of samples have already been taken in some haphazard fashion, but it is possible to supplement this array with some further sampling. This is one of the circumstances where some of the newer methods, broadly described as "kriging," may be especially useful.

There are various ways to go about making up a map from a given set of data. The major methods are summarized in the following sections.

4.5.2 Trend Surfaces

These are usually generated by the two-dimensional process analogous to curve-fitting by least squares. The most popular underlying model is that of a polynomial, and the major shortcoming of these polynomial fits is a tendency for strange behaviors at the edges of the fitted surfaces. Trend surfaces are most useful to give a broad general impression of the surface. The resulting fits may be used as components of the more elaborate stochastic process model (kriging) described below. Care in using polynomials is essential, since the higher degree models can yield some highly pathological results. Figure 4.5 suggests the kind of undesirable result that can be obtained, using data from the Beatty waste management site (Nielson, Wogman and Kirby 1980).

The various models for trend surface analysis may be subdivided as being linear and non-linear, and can be fitted to data by a variety of methods described in detail in some of the references given above. The best known approach is simply to use ordinary multiple regression obtaining an equation for a plane. As noted above, polynomials are frequently used to supply various kinds of curved surfaces. Since the powers and cross-products of the spatial coordinates are known (but are often selected arbitrarily), it is only the coefficients that must be estimated, so that the model is thus linear (in the unknown coefficients). Hence, multiple regression can again be used for the computations.

If x and y denote the spatial coordinates, and z the value of the surface (e.g., elevation above the plane of the coordinates, the simplest model is then:

FIGURE 4.5. A possible consequence of fitting trends with a polynomial model. The cesium-137 concentration in surface soil data (Z in the figure) modeled were taken from Table II (pp. 11-13) of Nielson, Wogman and Kirby (1980) and were obtained to define the area of radioactivity where waste canisters had been accidentally excavated. The polynomial fit described the observed data very well (not shown, but within the dark box in the figure). However, as can be seen above, the predicted concentrations very near the spill site (but beyond the observed data) are not reasonable (i.e. increase or decrease much too rapidly).

$$z = a + bx + cy$$

which is simply the equation of a plane surface in three dimensions. To get a curved surface, one might use:

$$z = a + bx + cy + dx^2 + ey^2 + fxy$$

Higher powers and more cross-products can be added so that almost any surface can be fitted. However, whether such a surface is realistic is another matter, as illustrated in Figure 4.5. The usual recommendation is to stay with the lower powers, i.e., quadratics (as in the above equation) or at most cubic equations. Even these simpler models can provide a good deal of flexibility. One can, for example, drop some of the linear terms (i.e., bx or cy) or the crossproduct term (fxy), and so on. One way to check on likely utility of trend surface models is to examine the "deviations from regression" obtained as successively more complex models are fitted to a given data set.

The non-linear models can be further subdivided as being "intrinsically linear" and "intrinsically non-linear," which simply means that there are non-linear models that can be converted to a linear form by a suitable transformation. If the model to be used is intrinsically non-linear, non-linear least-squares will be needed in fitting. An intrinsically linear model might be handled either way, i.e., transformed and fit by linear least-squares (i.e., multiple regression) or fit directly by non-linear least squares. The choice depends on what further assumptions are made about the underlying model (distribution of errors, correlations, etc.) and what will be done with the results. If the fitted surface is to be displayed or further analysis done in terms of the untransformed variables, a transformation may cause a number of problems. Probably the most frequent example of an intrinsically linear model in low-level waste management will be the various surfaces that might be constructed with exponential terms, since these are the most likely results from underlying physical principles and/or theory.

4.5.3 Moving Averages

An alternative approach is just to smooth the observations by using moving averages. Moving averages are constructed by replacing a given value by the average containing it and some neighboring points, starting at one end and moving through the series of points. Such schemes may not work at all well in the presence of clusters of points since the values in a cluster may exert too great an influence on points in sparsely populated surrounding areas. This shortcoming may be attacked by using some kind of weighting scheme so that the influence of a point depends on how distant it is from the location being mapped. Another useful refinement takes into account only the nearest points in a given quadrant.

4.5.4 Spatial Splines

In one dimension, curves may be fitted by using smooth curves to interpolate in the intervals between data points. Various constraints are put on these interpolatory curves near the data points to prevent erratic behaviors, i.e., the curves may be required to pass through the data points and should do so smoothly. In two dimensions, the problem is much more difficult since the convenient ordering of the points in one dimension is no longer available. We thus do not propose to investigate splines for the purposes of this study, unless some special need arises.

4.5.5 Stochastic Process Models

This approach is based on the assumption that the "true" surface being studied is only one example of an infinite number of possible outcomes from the process that generated it. This permits considering two components. One is an underlying large-scale trend, that would be present in any realization of the process. The other is a localized or small-scale fluctuation. In a waste management context, one may think of a process transporting radioactivity away from a storage site. Almost always, the process causing movement (e.g., wind or water movement) will have a strong directional component, so that one side of the area or volume will have much more movement than others. This may then be thought of as the trend component, on which there will be superposed a variety of smaller scale fluctuations resulting possibly from many causes.

It should be noted that trends and fluctuations often are only a matter of scale. If one considers small areas within a larger region, the fluctuations observed on a larger area may look like "trends" with smaller scale "fluctuations" of their own in the subarea.

A relevant model is:

$$Z(x,y) = m(x,y) + e(x,y)$$

where (x,y) denotes spatial coordinates of a sampling point, $Z(x,y)$ gives the value associated with a given point (e.g., concentration per gram of soil), $m(x,y)$ represents the trend and $e(x,y)$ denotes the random fluctuations at that point. A critically important point here is that $e(x,y)$ is not a random error in the sense of the usual statistical model. Observations taken a small distance apart will be very nearly the same, i.e., will be highly correlated. Both the difficulties and the advantages of this approach depend very much on the assumption of a correlation or covariance function to represent the similarity of nearby observations. In the waste management example suggested above, there may be measurement errors in determining the concentrations of radioactivity in each sample. These should not be confused with fluctuations as represented by $e(x,y)$.

With the model given above, one can elect to use an estimate of the trend, $m(x)$ to show overall behavior of the main feature of the process or use trend plus fluctuation as a way to interpolate between data points. There is thus more flexibility than is available with the previous methods. This comes at a price comprised of both complexity and additional assumptions. The simpler versions of the model assume that the covariance function and the trend are both known. Neither assumption is realistic in many practical problems.

Another difficulty is that the main approaches in use require the fitted functions to pass through the data points exactly. This is, of course, essential if it is assumed that the variable of interest is measured without error. However, in many waste management situations, this will not be true. For example, in sampling soil for radioactivity ordinarily only a small mass of soil can be assayed, so that the samples taken in the field are subsampled in the laboratory, and any two subsamples of the original sample cannot be expected to give the same result. One thus should not require that the fitted surface should pass through every data point.

4.5.6 Kriging

Kriging is essentially a method of estimation that assumes that the underlying model belongs to the class of stochastic process models described above. There are a number of variants of the basic method, and various special developments. Undoubtedly, more variations on the basic themes will soon surface, and others will have escaped our attention. The existence of the several varieties of kriging, along with the somewhat elaborate mathematics involved in the original development (initially published mainly in French), have led to confusion.

A comment by M. David (cited in full below in the review of the book by Guarascio, David and Huijbregts 1976) that "kriging...can be expressed as plain regression" is quite correct for some elementary versions. A linear model is assumed, and the main difference from ordinary multiple regression is that a weighting is used in consequence of correlations among variables. Weighted multiple regression is, of course, not a new development. The difficulties come in at several points, which we only briefly touch on here. A rather more detailed assessment is needed to ascertain the ways in which the methodology can be useful in low-level waste management.

Perhaps the most essential issue that needs to be addressed is whether the right question is being asked about the data. Mining engineers and petroleum geologists are normally concerned with estimating a quantity of some kind, usually that of the economically recoverable resource in a region. The appropriate methodology is thus one aimed at efficiently estimating a total or a mean. In low-level waste management, the key questions are very often not concerned with totals or averages. It is thus quite possible that good answers to the wrong questions may be provided by a particular methodology. Hence, we believe that it is essential to look quite carefully at the several versions of kriging

(and related methodology) before recommending their use in low-level waste management contexts.

A second issue concerns the underlying model assumptions made in a given application. These assumptions need to be considered in several stages. A first, and fundamental, stage is the choice of the "process" model. The equation given in Section 4.5.5 can conveniently be illustrated by contemplating making a topographic map. One goes out to a series of locations ("control points") and accurately determines elevations. In many situations, the topography is such that there will be a "trend" (i.e., the general "lay of the land"), on which various local "fluctuations" are superimposed (hills, streams, etc.). If one constructs a contour map from such data, the contours on the map should coincide exactly with the observed data points. As has been remarked previously here, this is not necessarily true in the waste management context. For example, soil sampling for radioactivity will seldom give the same result at adjacent points, or for repeated samples from any very small area. Consequently, the underlying model will often have three components:

$$Z(x,y) = m(x,y) + f(x,y) + e(x,y)$$

where $m(x,y)$ again represents the trend, $f(x,y)$ represents a fluctuation (i.e., a real local variation in concentration), and $e(x,y)$ is now an "error" term in the usual sense. The fitted model will not, and should not, coincide with the observed data points. A computer program that forces such a fit may be quite inappropriate for the data.

A realistic application to low-level waste management thus may require three components in the underlying model -- an overall trend, identifiable local fluctuations, and an error component. An immediate difficulty is that it is well-established that radionuclide concentrations almost invariably follow a skewed frequency distribution. Since virtually all of the theory for estimation methodology (e.g., kriging) assumes normally distributed errors, a transformation of the data is required. Usually, in both geology and low-level waste management, a logarithmic transformation is appropriate. As was noted in discussing trend surfaces above, a non-linear model may be needed to account for the trend, and may possibly be transformable to a linear model, so that two situations calling for transformations may be encountered. There can be no assurance that they will be mutually compatible.

The next issue requiring attention is the correlation between nearby locations in the region being studied. A basic assumption in kriging is the so-called "weak stationarity." Two aspects need to be considered. The first is "first order" stationarity, which assumes that the mean value (average) of the variable of interest is constant over the region being studied. This is usually stated by considering a (vector) distance (h) between two points and indicating that the expected value of the differences is zero, i.e., that:

$$E[Z(x,y) - Z((x,y)+h)] = 0$$

Since this will not be true in many practical cases, we may have to postulate the trend component, $m(x,y)$, of the model above and arrange to "remove" a trend to obtain first order stationarity.

A second component involves a "second order" relationship, which is represented by the expectation of the squared difference of the variable of interest measured at two locations. This is the "intrinsic hypothesis" of kriging:

$$E[(Z(x,y) - Z((x,y)+h))^2] = 2g(h)$$

i.e., that the expectation of squared differences depends only on a function of the distance between two locations. Plots of this function, $g(h)$, are called the "semivariogram" or sometimes, just "variogram". There is a simple relationship to a correlogram, showing the correlations between values recorded at points a distance h apart.

The expression above can conveniently be thought of as a "spatial variance", i.e., it measures the variance between variables that are associated with definite locations. These are Matheron's "regionalized variables" (in the usual statistical theory, "random variables" are used in abstraction from any physical location).

To use kriging, one must arrive at a functional form for $g(h)$, and then estimate the constants involved in that functional form. An exponential model is frequently assumed, and one of the constants will often represent the error term, $e(x,y)$ above (actually, the error variance). In applications, it may be difficult to obtain enough suitable data to actually fit a variogram to data from the region under study. A poorly estimated or arbitrary function may thus be involved in the estimation process.

In a final step in the process, a series of weights (adding to unity) are obtained through a Lagrangian multiplier operation on a system of equations based on variogram values for the observed locations. These weights then can be used in a moving average process to generate values of the variable of interest over the entire study region, along with variances.

It seems quite evident that there are a number of places where the estimation process can go astray, in the sense of uncertainty as to how well the underlying assumptions are met on real data. We expect that an essential step is to look rather carefully at the needs for kriging or related processes in low-level waste management before attempting to probe deeper into the methodology of the several variants of the method. In those instances where field data are not extremely expensive, it may be much better to collect some additional data, rather than to attempt to use a complex process on inadequate data. When additional

observations are very expensive or impossible to obtain, then the effort required to find the "best" methodology may be justified.

4.5.7 Some Applications of Kriging

A number of applications of "geostatistics" (which is sometimes taken to mean kriging, although it has a wider meaning) have been made in a context relevant to low-level waste management. A few will be mentioned here.

Doctor (1980) used kriging to study the top of the basalt formation at the Hanford Site in eastern Washington. This example is much like the exercise in topographic mapping used to illustrate concepts above, in that the contributions of measurement errors to the error term, $e(x,y)$, are inconsequential. Two techniques not discussed in Section 4.5.6 were used, one being the use of a "generalized covariance analysis" to select a drift and covariance structure, and the other partitioning of the area into subareas to account for local features that make an overall analysis difficult. An important suggestion is included, i.e., that kriging may provide a basis for selecting new locations for collecting data to improve the accuracy of a contour map made from existing data. We believe that this may be one of the major potential uses in low-level waste management. In an earlier study (Doctor 1979), ground-water levels were assessed by kriging, using data sets for two different years on the same set of wells. Since a large set of wells was available, there is detail enough in some sub-regions for various possible cross-checks on the method. That is, some authors have recommended carrying out the various available methods on subsets of data, and predicting the observations at sampling locations for which an independent set of data exists, as a check on validity of the kriging approach.

Another application (Devary and Doctor 1981) deals with waste management aspects of a geologic formation. In this instance, media conductivity and ground-water potential gradients were assessed. It was necessary to take into account a stochastic error, which might be a measurement error, or reflect the fact that an on-going process (ground water flow, for example) is variable over time. This study thus provides some different angles on phenomena and methodology, all relevant to waste management studies.

Two other studies (White, Simpson and Bostick 1980, and White 1981) utilize kriging to assess surface contaminations of uranium, thus having a potential similarity to waste management needs. These studies are particularly interesting in that they were based on data collected on a polar coordinate scheme, to intensify sampling near the origin. In cases where a point source can be identified for contamination, such a sampling scheme may have special advantages. Experience in the first study led to a modification of the plan in the second study, in which some supplemental sites were interspersed between the rays of the polar scheme. Data from both of these investigations might be worth further study, by way of investigating the question of the structure of the model assumed. In both cases, measurement errors probably were present,

but appear to have been lumped with the "fluctuation" component of the kriging model.

In nearly all of the kriging references we have examined, the trend is either assumed to be linear, or modelled by a polynomial. A study by Volpi and Gambolati (1978) is thus of some interest, since they investigated the use of polynomials, and obtained some anomalous results. They recommend, in cases where a linear trend is not practical, that an analytic trend model should be determined in advance. They also point up the problems in determining the appropriate variogram and supporting that choice by reference to actual data. This difficulty is apparent in most of the references describing actual applications, leading to a common view of the techniques as an empirical approach. Gambolati and Volpi (1979) describe an application of their views to ground-water contour mapping in Venice, where kriging was used to study the change in the groundwater surface after reduction in the drawdowns in a particular industrial area. Utilization of kriging for designing a ground-water network has been discussed by Sophocleous, Paschetto and Olea (1982), and the approach and results should be of interest in planning similar work with respect to waste management sites.

4.5.8 References on Sampling for Spatial Patterns

There is a sizable and rather scattered literature on sampling for pattern. Much of the original work on "kriging" is in geology and mining, and a good deal of the relevant early literature was published in France and South Africa. The "French school" (Matheron and his colleagues and students) have continued to produce substantial volumes of contributions, not all of which is readily accessible. The other main field of application and of development of theory is in geography. Probably the developments there will be of less interest in the present context, but there are undoubtedly various aspects of value in low-level waste management. A more detailed assessment of the literature is needed in the near future. For the present, we list some books that provide convenient references for several aspects of theory and practice. Several other books should be included, but have not been as yet reviewed in sufficient detail.

Agterberg, F. P. 1974. Geomathematics. xvi + 596 pp. Elsevier Scientific Publishing Co., Amsterdam, London, and New York.

This is a rather ambitious textbook, aimed at providing mathematical and statistical methods for geologists. Chapters are devoted to the calculus, matrix methods, geometry, factor analysis, probability and statistics, frequency distributions (and functions of random variables), and multiple regression. These serve as a prelude to coverage of trend analysis, kriging, harmonic analysis, and Markov chains, for all of which illustrations on geological data are provided. In the present context, the main interest is in the material on trend analysis and kriging. These chapters (Ch. 9 and 10) should be useful to anyone wishing to review the basis for geological use of the several methods, but will probably not be adequate as a single reference.

 Cliff, A. D., P. Haggett, J. K. Ord, K. A. Bassett and R. B. Davies.
 1975. Elements of spatial structure. xvii + 258. Cambridge Univ.
 Press, London, New York, Melbourne.

This book is essentially concerned with geography, or at least with data handled on geographic regions. For most of the low-level waste management needs, it will probably be preferable to use references from geology or related fields, since the variables are often much more similar. In particular, concentrations are of substantial importance in geological applications, and the log-normal distribution characteristic of concentrations of radionuclides is also widely used in geology. This is not to say that the book mentioned above is not relevant, since many of the techniques are comparable to those seen elsewhere. However, its main utility would seem to be in a research sense, considering techniques that might be adapted for special problems in waste management, and not as a primary reference.

 Cliff, A. K., and J. K. Ord. 1973. Spatial autocorrelation. x + 178.
 Pion Ltd., London.

This is a technical reference, concerned in large part with testing for autocorrelation. It is also mainly concerned with geography, and thus of less interest here than the geological references. The material covered will, however, be useful in examining particular problems in research on low-level waste management needs. In most instances, we likely will not be especially interested in testing for autocorrelation directly, but the tests may be of considerable utility in examining the applicability of particular theoretical models of autocorrelation to spatial data collected in waste management studies. Another feature of some importance is that there is a chapter on the analysis of regression residuals for autocorrelation. This is important in that it is desirable to attempt to determine whether or not the models fitted to low-level waste data do in fact provide a satisfactory fit. Usually the reliability of the fit has to be tested by examining residuals, i.e., deviations between observations and model predictions. If the model assumed for autocorrelation is satisfactory, then the residuals should not be correlated. Another related issue is that "stationarity" is assumed in several of the more widely applied techniques for dealing with spatial data, and some of the work discussed by Cliff and Ord will be useful in examining the validity of this assumption.

 Davis, J. C. 1973. Statistics and data analysis in geology. xiii + 550. J. Wiley and Sons, New York.

This is an elementary book aiming to supply background for "earth scientists" in the applications of basic mathematics and statistics to geological and related problems. There are chapters on computers, elementary statistical methods, matrix algebra, time series, analysis of maps, and multivariate analysis. FORTRAN programs are provided for many of the techniques covered, thus providing students with an opportunity

to try out the basic methods. This book is rather more elementary than that by Agterberg (described above), since calculus is largely avoided. It will, however, be a useful reference to anyone wanting a basic description of the methods, so long as it is supplemented by some more detailed references.

Davis, J. C., and M. J. McCullagh, Editors. 1975. Display and analysis of spatial data. xiv + 378. J. Wiley and Sons, New York.

This is a collection of papers resulting from a NATO Advanced Study Institute, held in England in 1973. It is a particularly valuable reference due to the diversity of material covered. Unlike many of the other references, it is not dedicated solely to kriging and related approaches, but does, of course, have much material on that subject. There is a good deal of material on mapping, some of which is not particularly relevant to the low-level waste management situation. Since the collection of papers is quite diverse, we will not attempt to cover it here, but will instead plan to use the individual papers in the research projected in the next stage of this program.

Guarascio, M., M. David, and C. Huijbregts, Editors. 1975. Advanced geostatistics in the mining industry. xvi + 461. D. Reidel Publ. Co., Dordrecht, Holland, Boston.

This book contains the proceedings of a symposium (NATO Advanced Study Institute) held in 1975. It covers several subjects of interest in the present context. Unfortunately, the two papers covering basic concepts are in French (both are by Matheron, and are largely expository). Five succeeding papers are on kriging. The first of these (by M. David) has an interesting comment on kriging: "As a technique in ore reserve estimation, kriging as we now know it is already 15 years old, and still puzzling most of the mining people. As a statistical technique, it has nothing new, since it can be expressed as plain regression with or without correlated residuals (Watson 1971), as such it has been briefly tested in simple problems, most of the time in two dimensions, and generally found to only have a marginal advantage in the quality of the estimate, and a definite drawback in its computer cost. There is no doubt that these comparative exercises are correct, for the particular situations tested, and with the programs used. It is also true that if a brute force method is applied, simply following the standard kriging algorithm, the computing cost will be more than for other methods." Unfortunately, the reference to Watson (1971) was left out of the reference list, but it is most likely a paper by G. S. Watson in the Journal of Mathematical Geology.

The balance of the section on kriging contains four papers that cover several aspects of the techniques as they apply to mining geology and engineering. Five subsequent sections deal with both applications and theory. Probably the main features of this collection can be summarized by saying that the emphasis is mainly on kriging and on applications to mining. There are thus a number of useful results needing further

consideration, but this needs to be done with the fact in mind that the major application is one of estimating totals or means, i.e., the value of mineral deposits. These purposes are sufficiently different from those of low-level waste management that considerable caution needs to be exercised in attempting any transfer of methodology and theory. An interesting paper in this collection is by D. G. Krige, whose name was attached to the techniques pioneered by G. Matheron. Krige gives the history and reviews development of the original work with respect to its use in the South African mining industry (primarily gold-mining).

Henley, S. 1981. Nonparametric geostatistics. xiv + 145 pp. J. Wiley and Sons, New York.

This short book should be quite useful in the present context, due to its abbreviated, but reasonably clear, discussions of the many techniques in use for treating spatial patterns in geology. The main purpose of the book is to propose an alternative to kriging in the form of a distribution-free or "non-parametric" approach. Unfortunately, we have not been able to go into the methods suggested in enough detail to offer much of a critique here.

Lewis, P. 1977. Maps and statistics. xviii + 318. J. Wiley and Sons, New York.

This book is primarily concerned with the analysis of maps and mapped data using basic statistical and probability methods. As such it may not be of much direct interest in low-level waste studies, except for research on particular problems having to do with the spatial pattern of objects and observations. Where data have been collected in a more or less haphazard manner, so that utilizing the results is something of a gamble, with respect to possible selectivity, gaps, and so on, it may be useful to study the pattern of locations for clues as to possible difficulties. In this case, methodology for assessing maps is appropriate and useful. A variety of parametric and non-parametric tests are covered in the book.

Merriam, D. F., Editor. 1970. Geostatistics. xiii+177. Plenum Press, New York.

This is a collection of papers given at a colloquium at the University of Kansas in 1970. They are generally concerned with mathematical issues of concern in "geostatistics," but there is no particular theme. The material may thus be of interest and utility in research on low-level waste problems, but cannot be recommended for a particular use. The collection contains several papers with relevance to kriging, including a short exposition of the methodology by Matheron.

Ripley, B. D. 1981. Spatial statistics. ix+252. J. Wiley and Sons, New York.

This short book attempts to cover much of "spatial statistics," and provides some valuable and concise summaries of many of the methods now in use. At several points, the presentation is too brief for ready comprehension, unless the reader has a good background in mathematical statistics and some familiarity with the material being discussed. There are, however, enough good, brief and non-mathematical explanations to make it a useful reference for anyone interested in the topic.

5.0 SUMMARY AND CONCLUSIONS

The overall objective of the project under which this report was written is to do the research necessary to provide guidance for the use of statistics in planning, conducting surveys, monitoring and assessing data relative to the siting, operation and decommissioning of low-level waste disposal sites. In the Statement of Work, which initiated the project, the Nuclear Regulatory Commission staff asked for "an initial assessment of the substantial theoretical information and experience that may be adapted for NRC needs." The present report seeks to provide that assessment. Since it is an initial assessment, we will not attempt to reach any final conclusions here, beyond re-emphasizing some of the features that we believe need to be adapted for NRC needs. We note that "NRC needs" are very diverse, ranging from the specifics of closing particular existing sites (such as Maxey Flats or Sheffield) to the generic issues of providing guidance for selecting, operating, and closing sites in the future. Statistical and sampling requirements will undoubtedly differ greatly in the two situations, but hopefully dealing with present problems will yield useful insights for planning for the future.

In the present report, we have described four classes of sampling (descriptive, analytical, pattern and sampling for model parameters) and attached convenient labels to each approach. Taken together, they can encompass much of modern statistical methodology, so we do not suppose that a single report can cover the field. The major point that we would like to emphasize as a conclusion is just that the efficient choice of methods depends very much on the primary objectives of the study. By "efficient choice" we mean cost-effectiveness -- getting the best results for dollars spent. Our experience with practical sampling problems in the nuclear field has been that it is often not possible to define objectives very narrowly. Nonetheless, choosing a particular sampling scheme does amount to choosing a particular kind of objective, whether those who make the choice realize it or not.

A stratified random sample with optimum allocation is almost always the best way to estimate total quantities (or average concentrations). It is, however, generally a very poor approach if it is the pattern of distribution that is of primary interest. In the example presented in the Introduction to Chapter 2.0, choosing a plan for estimating a total when pattern is really of main interest amounts to wasting roughly two thirds of the sample. Under the usual conditions encountered in assessing radionuclides in a field setting, a sampling plan aimed at making comparisons between subregions of an area being studied may take an intermediate course between these two extremes (sampling for a total or sampling for pattern). As we have thus far defined it, sampling for modelling tends to stand apart from the other three classes of sampling, since it concentrates on parameter estimation for defined models applied in one dimension. We believe, however, that an adequate understanding of the theoretical basis may serve to improve applications of sampling for pattern.

Our main recommendation to decision-makers at this juncture in the study is that they try very hard to narrow down the objectives for a particular study before a plan is laid out. If the objectives nonetheless turn out to be quite broad, then we recommend that "optimal" sampling plans for each of the four classes of sampling should be determined and compared. This will often help narrow down the objectives, and selection of a particular pattern, compromise though it might be, will leave all concerned with a better understanding of what they are likely to accomplish, and have decided to do.

The balance of this summary lists some salient needs developed in the main body of the report, and attempts to indicate what research may be needed to provide better sampling tools for low-level waste management.

1. An assessment of the accuracy of standard textbook assumptions for use of key statistical methods in low-level waste management situations (e.g., normal and additive errors) is needed, followed by recommendations concerning the use of transformations (Section 2.3).
2. Detailed guidelines on compositing and pooling in the various circumstances where these techniques may be applied should be developed. Data from various field sampling applications need to be assembled and analyzed for this purpose. Adequate (replicated) data may not be available, so that specific experiments may well be required to develop detailed guidance. A simulation study of compositing for detection is underway, and may need to be extended. Recommendations on the censoring problem ("below detection limits") should be included (Sections 3.1 through 3.6).
3. Guidelines are needed for the sequential use of the four classes of sampling. This should include ways to use comparisons of sampling plans devised for each class of methods to help define a set of objectives for a given study (Section 4.1).
4. An evaluation of appropriate circumstances for use of systematic sampling and of the hazards attendant on its use when the statistical method is based on random sampling is needed (Section 4.2.2).
5. More detailed recommendations on use of stratified sampling and a review and evaluation of cost functions and allocation schemes should be developed (Section 4.2.4).
6. A simulation study of double-sampling should be done to refine the present rules of thumb as to minimal feasible sample sizes under various circumstances (Section 4.2.5).
7. A further analysis of requirements for repeated sampling over time is needed, with an elaboration of existing rules as they apply to low-level waste sites (Section 4.2.6).
8. Guidance on sample size requirements for particular circumstances should be developed (Section 4.2.7).

9. The recommendations in the present report as to a choice of references on descriptive (survey) sampling should be refined (Section 4.2.8).
10. Research to provide more definite guidelines for analytical sampling should be conducted; in particular to define circumstances in which bona fide (or approximately so) replicate samples can be considered to be available. More details on multiple comparison methods for low-level waste management uses are needed. An assessment of the relative weights that might be assigned to the set of assumptions for statistical analyses should be made, i.e., when they conflict (as is often the case), which is likely to be critical (Sections 4.3.2, 4.3.3 and 4.3.5)?
11. Further consideration of the relevance of the "superpopulation" approach for devising sampling schemes in low-level waste management is needed (Section 4.3.5).
12. A specific list of optimal times for a variety of models likely to be useful in field circumstances should be provided, supported by computer routines for calculating optimal times when an analytical solution is difficult or not available (Section 4.4.2).
13. Both the concepts of replication and contrasting models should be utilized to provide schemes for extending the basic set of optimal times to a more practical working approach (Section 4.4.5).
14. A framework for applications of sampling for modelling to low-level waste site needs is needed; in particular the interrelationships with sampling for pattern should be explored (Section 4.4.6).
15. The recent literature on the Box and Lucas scheme should be reviewed for selection of applicable references in low-level waste contexts (Section 4.4.7).
16. More specific recommendations on use of sampling for pattern are needed, especially in distinguishing between situations where existing data must be used (possibly supplemented by selection of additional new samples) and those where a new sampling plan can be utilized (Section 4.5.1).
17. The circumstances where kriging is and is not appropriate for low-level waste management use should be defined and situations where less complicated methods may be preferred should be indicated. This will require further study of the impacts of assumptions about variograms and trend surfaces, which will likely require computer simulations and may also call for some field sampling experiments (Section 4.5.6).
18. A further evaluation of available references on kriging and geostatistics and an extension of present reference list and review is needed (Section 4.5.8).

6.0 REFERENCES CITED

- Agtergerg, F. P. 1974. Geomathematics. Elsevier Scientific Publishing Co., Amsterdam, London and New York. xvi + 596 pp.
- Atkinson, A. C. 1969. "A Test for Discriminating Between Models." Biometrika 56:337-347.
- Atkinson, A. C. 1972. "Planning Experiments to Detect Inadequate Regression Models." Biometrika 59:275-293.
- Atkinson, A. C., and W. G. Hunter. 1968. "The Design of Experiments for Parameter Estimation." Technometrics 10:271-289.
- Bailar, B. A., and C. M. Lanphier. 1978. Development of Survey Methods to Assess Survey Practices. American Statistical Association, Washington, D.C. x + 117 pp.
- Bard, Y. 1974. Nonlinear Parameter Estimation. Academic Press, New York. x + 341 pp.
- Beck, J. V., and K. J. Arnold. 1977. Parameter Estimation in Engineering and Science. John Wiley and Sons, New York. xix + 501 pp.
- Box, G. E. P., and W. J. Hill. 1967. "Discrimination among Mechanistics Models." Technometrics 9:57-71.
- Box, G. E. P., W. G. Hunter and J. S. Hunter. 1978. Statistics for Experimenters. John Wiley and Sons, New York. xviii + 653 pp.
- Box, G. E. P., and G. M. Jenkins. 1970. Time Series Analysis, Forecasting and Control. Holden-Day, San Francisco, California. xix + 553 pp.
- Box, G. E. P., and H. L. Lucas. 1959. "Design of Experiments in Non-Linear Situations." Biometrika 46:77-90.
- Box, M. J. 1970. "Some Experiences With a Nonlinear Experimental Design Criterion." Technometrics 12:569-589.
- Box, M. J. 1971a. "An Experimental Design Criterion for Precise Estimation of a Subset of the Parameters in a Nonlinear Model." Biometrika 58:149-153.
- Box, M. J. 1971b. "Simplified Experimental Design." Technometrics 13:19-31.
- Brown, G. H., and N. I. Fisher. 1972. "Subsampling a Mixture of Sampled Material." Technometrics 14(3):663-668.

- Campbell, D. T., and J. C. Stanley. 1966. Experimental and Quasi-Experimental Designs for Research. Houghton Mifflin Co., Boston, Maine. ix + 84 pp.
- Cassel, C., C. Sarndal and J. H. Wretman. 1977. Foundations of Inference in Survey Sampling. John Wiley and Sons, New York. vii + 192 pp.
- Cherry, J. A., P. Fritz, J. E. Gale and R. W. Gilham. 1979. "Evaluation Methods for Hydrogeologic Conditions at Radioactive Waste Burial Sites." Trans. Amer. Nuclear Soc. 32:117-119.
- Cleary, J. P., and H. Levenbach. 1982. The Professional Forecaster: The Forecasting Process Through Data Analysis. Lifetime Learning Publications, Belmont, California. xi + 402 pp.
- Cliff, A. D., P. Haggett, J. K. Ord, K. A. Bassett and R. B. Davies. 1975. Elements of Spatial Structure. Cambridge University Press, London, New York and Melbourne. xvii + 258 pp.
- Cliff, A. K., and J. K. Ord. 1973. Spatial Autocorrelation. Pion Ltd., London. x + 178 pp.
- Cochran, W. G. 1973. "Experiments for Nonlinear Functions." J. Amer. Stat. Assoc. 68:771-781.
- Cochran, W. G. 1977. Sampling Techniques. 3rd ed. John Wiley and Sons, New York. xvi + 428 pp.
- Code of Federal Regulations, Part 61 - Licensing Requirements for Land Disposal of Radioactive Wastes, 47 Fed. Reg. 57463-57477 (December 7, 1982).
- Commonwealth of Kentucky, Natural Resources and Environmental Protection Cabinet. 1982. Fifth Quarterly Report for the Period June 1-August 31, 1982. Appendix 5, Specific Analysis of Surface Stream Sampling. Submitted to U.S. Nuclear Regulatory Commission, Washington, D.C. 20555. Copies available from: Maxey Flats Branch, Dept. Environmental Protection Cabinet, 18 Reilly Road, Ft. Boone, Frankfort, Kentucky 40601.
- Commonwealth of Kentucky, Natural Resources and Environmental Protection Cabinet. 1982. Fifth Quarterly Report for the Period June 1-August 31, 1982. Appendix 6, Maxey Flats Surface Water Hydrology Study. Submitted to U.S. Nuclear Regulatory Commission, Washington, D.C. 20555. Copies available from: Maxey Flats Branch, Dept. Environmental Protection Cabinet, 18 Reilly Road, Ft. Boone Plaza, Frankfort, Kentucky 40601.
- Davis, J. C. 1973. Statistics and Data Analysis in Geology. John Wiley and Sons, New York. xiii + 550 pp.

- Davis, J. C., and M. J. McCullagh, eds. 1975. Display and Analysis of Spatial Data. John Wiley and Sons, New York. xiv + 378 pp.
- Deming, W. E. 1943. Statistical Adjustment of Data. John Wiley and Sons, New York. (Available in a Dover reprint, 1964). x + 257 pp.
- Deming, W. E. 1950. Some Theory of Sampling. John Wiley and Sons, New York. (Available in a Dover reprint, 1966).
- Deming, W. E. 1960. Sample Design in Business Research. John Wiley and Sons, New York.
- Devary, J. L., and P. G. Doctor. 1981. Geostatistical Modeling of Pore Velocity. PNL-SA-9096, Pacific Northwest Laboratory, Richland, Washington. 33 pp.
- Dixon, D. O., and D. B. Duncan. 1975. "Minimum Bayes Risk T-Intervals for Multiple Comparisons." J. Amer. Statist. Assoc. 70:822-831.
- Doctor, P. G., and R. O. Gilbert. 1978. Two Studies in Variability for Soil Concentrations: With Aliquot Size and With Distance. PNL-SA-6783, REV., NVO-192. Pacific Northwest Laboratory, Richland, Washington. 43 pp.
- Doctor, P. G. 1979. An Evaluation of Kriging Techniques for High Level Radioactive Waste Repository Site Characterization. PNL-2903, Pacific Northwest Laboratory, Richland, Washington. 67 pp.
- Doctor, P. G. 1980. "The Use of Geostatistics in High Level Radioactive Waste Repository Site Characterization." Radioactive Waste Manage. 1(2):193-210.
- Dorfman, R. 1943. "The Detection of Defective Members of Large Populations." Annals Math. Stat. 14:436-440.
- Duncan, A. J. 1962. "Bulk Sampling: Problems and Lines of Attack." Technometrics 4:319-344.
- Duncan, D. B. 1955. "Multiple Range and Multiple F Tests." Biometrics 11:1-42.
- Dunnett, C. W. 1980a. "Pairwise Multiple Comparisons in the Homogeneous Variance, Unequal Sample Size Case." J. Amer. Stat. Assoc. 75:789-795.
- Dunnett, C. W. 1980b. "Pairwise Multiple Comparisons in the Unequal Variance Case." J. Amer. Stat. Assoc. 75:796-800.
- Eberhardt, L. L. 1975. "Sampling for Radionuclides and Other Trace Substances." In Radioecology and Energy Resources Fourth National Radioecology Symposium, ed. C. E. Cushing, Jr., pp. 199-208. Dowden, Hutchinson and Ross, Inc., Stroudsburg, Pennsylvania.

- Eberhardt, L. L. 1976. "Quantitative Ecology and Impact Assessment." J. Environ. Manage. 4:27-70.
- Eberhardt, L. L. 1978. "Designing Ecological Studies of Trace Substances." In Proceedings of Symposium on Environmental Chemistry and Cycling Processes, eds. D.C. Adriano and I. L. Brisbin, Jr., pp. 8-33. CONF-760429, National Technical Information Service, Springfield, Virginia
- Eberhardt, L. L., and R. O. Gilbert. 1976. "Sampling the Environs for Contamination." In Proceedings of the First ERDA Statistical Symposium, pp. 187-208. Los Alamos Scientific Laboratory, BNWL-1986, Battelle, Pacific Northwest Laboratory, Richland, Washington.
- Eberhardt, L. L., and R. O. Gilbert. 1980. "Statistics and Sampling in Transuranic Studies." In Transuranic Elements in the Environment, ed. W. C. Hanson, pp. 173-186. DOE/TIC-22800, National Technical Information Service, Springfield, Virginia.
- Eberhardt, L. L., R. O. Gilbert, H. L. Hollister and J. M. Thomas. 1976. "Sampling for Contaminants in Ecological Systems." Environ. Sci. Technol. 10(9):917-925.
- Elder, R. S., W. O. Thompson and R. H. Myers. 1980. "Properties of Composite Sampling Procedures." Technometrics 22(2):179-185.
- Endrenyi, L., ed. 1981. Kinetic Data Analysis. Plenum Press, New York. ix + 427 pp.
- Gambolati, G., and G. Volpi. 1979. "Groundwater Contour Mapping in Venice by Stochastic Interpolators. I. Theory." Water Resources Res. 15(2):281-290.
- Gilbert, R. O., and L. L. Eberhardt. 1976. "An Evaluation of Double Sampling for Estimating Plutonium Inventory of Surface Soil." In Radioecology and Energy Resources Fourth National Radioecology Symposium, ed. C. E. Cushing, Jr., pp. 157-163. Dowden, Hutchinson and Ross, Inc., Stroudsburg, Pennsylvania.
- Gilbert, R. O. 1978. "Some Statistical Aspects of Compositing Field Samples. TRAN-STAT: Statistics for Environmental Transuranic Studies, Number 3." PNL-SA-6697, Pacific Northwest Laboratory, Richland, Washington. 21 pp.
- Gilbert, R. O. 1982. "Some Statistical Aspects of Finding Hot Spots and Buried Radioactivity. TRAN-STAT: Statistics for Environmental Studies, Number 19." PNL-SA-10274, Pacific Northwest Laboratory, Richland, Washington. 26 pp.

- Gilbert, R. O., and R. R. Kinnison. 1981. "Statistical Methods for Estimating the Mean and Variance from Radionuclide Data Sets Containing Negative, Unreported, or Less Than Values." Health Phys. 40:377-390.
- Green, R. H. 1979. Sampling Design and Statistical Methods for Environmental Biologists. John Wiley and Sons, New York. xi + 257 pp.
- Guarascio, M., M. David and C. Huijbregts, eds. 1976. Advanced Geostatistics in the Mining Industry. D. Reidel Publishing Co., Boston, Maine. xvi + 461 pp.
- Hajek, J. 1981. Sampling From a Finite Population. Marcel Dekker, Inc., New York. v + 247 pp.
- Hansen, M. H., W. N. Hurwitz and W. G. Madow. 1953. Sample Survey Methods and Theory. Volume 1. Methods and Applications. John Wiley and Sons, New York. xxii + 638 pp.
- Henley, S. 1981. Nonparametric Geostatistics. John Wiley and Sons, New York. xiv + 145 pp.
- Hill, W. J., and W. G. Hunter. 1966. "A Review of Response Surface Methodology: A Literature Survey." Technometrics 8:571-590.
- Hill, W. J., and W. G. Hunter. 1974. "Design of Experiments for Subsets of Parameters." Technometrics 16:425-434.
- Hill, W. J., W. G. Hunter and D. W. Wichern. 1968. "A Joint Design Criterion for the Dual Problem of Model Discrimination and Parameter Estimation." Technometrics 10:145-160.
- Hochberg, Y., G. Weiss and S. Hart. 1982. "On Graphical Procedures for Multiple Comparisons." J. Amer. Stat. Assoc. 77:767-772.
- Hunter, W. G., and A. M. Reiner. 1965. "Designs for Discriminating Between Two Rivals Models." Technometrics 7:307-323.
- Hwang, F. K. 1976. "An Optimum Nested Procedure in Binomial Group Testing." Biometrics 43:939-943.
- Jessen, R. J. 1978. Statistical Survey Techniques. John Wiley and Sons, New York. vi + 520 pp.
- Kane, V. E. 1979. Fitting a Three-Parameter Lognormal Distribution with Applications to Hydrogeochemical Data from the Natural Uranium Resource Evaluation Program. Oak Ridge Gaseous Diffusion Plant, (Union Carbide Corporation) Rept. K/UR-27, Oak Ridge, Tennessee
- Kish, L. 1965. Survey Sampling. John Wiley and Sons, New York. ix + 643 pp.

- Kramer, C. Y. 1956. "Extension of Multiple Range Tests to Group Means with Unequal Numbers of Replications." Biometrics 12:307-310.
- Kusssmaul, K., and R. L. Anderson. 1967. "Estimation of Variance Components in Two-Stage Nested Designs with Composite Samples." Technometrics 9:373-389.
- Latta, R. B. 1981. "A Monte Carlo Study of Some Two-Sample Rank Tests With Censored Data." J. Amer. Stat. Assoc. 76(375):713-719.
- Levenbach, H., and J. P. Cleary. 1981. The Beginning Forecaster: The Forecasting Process Through Data Analysis. Lifetime Learning Publications, Belmont, California. vi + 372 pp.
- Lewis, P. 1977. Maps and Statistics. John Wiley and Sons, New York. xviii + 318 pp.
- Lu, A. H. 1978. "Modeling of Radionuclide Migration from a Low-Level Radioactive Waste Burial Site." Health Phys. 34:39-44.
- McKenzie, D. H., L. L. Cadwell, C. E. Cushing, Jr., et al. 1982. Relevance of Biotic Pathways to the Long-Term Regulation of Nuclear Waste Disposal. A Report on Tasks 1 and 2 of Phase 1. NUREG/CR-2675, PNL-4241, Vol. 1, U. S. Nuclear Regulatory Commission, Washington, D.C.
- Mead, R., and D. J. Pike. 1975. "A Review of Response Surface Methodology from a Biometric Viewpoint." Biometrics 31:803-851.
- Mendenhall, W., L. Ott and R. L. Scheaffer. 1971. Elementary Survey Sampling. Wadsworth Publishing Co., Belmont, California. vi + 247 pp.
- Merriam, D. F., ed. 1970. Geostatistics. Plenum Press, New York. xiii + 177 pp.
- Narasimhan, T. N. 1979. "Groundwater Modeling: Industrial Problems and Regulations." Amer. Nucl. Soc. Trans. 32:116-120.
- Newman, D. 1939. "The Distribution of Range in Samples from a Normal Population, Expressed in Terms of an Independent Estimate of Standard Deviation." Biometrika 31:20-30.
- Nielson, H. L., N. A. Wogman and L. J. Kirby. 1980. Field Survey of the Shallow Land Low-Level Radioactive Waste Burial Site Near Beatty, Nevada. NUREG/CR-1534, PNL-3427, U. S. Nuclear Regulatory Commission, Washington, D.C. 33 pp.
- Nuclear Regulatory Commission Staff. 1974. Measurements of Radionuclides in the Environment. Regulatory Guide 4.5. U. S. Nuclear Regulatory Commission, Washington, D.C. 9 pp.

- Nyhan, J. W., T. J. Schofield, G. C. White and G. Truhillo. 1981. Sampling Soils for 137-Cs Using Various Field Sampling Volumes. LA-8951-MS, Los Alamos National Laboratory, Los Alamos, New Mexico. 12 pp.
- Phillips, S. J., L. L. Ames, R. E. Fitzner, G. W. Gee, G. A. Sandness and C. S. Simmons. 1980. Characterization of the Hanford 300 Area Burial Grounds. Final Report. Decontamination and Decommissioning. PNL-2557, Pacific Northwest Laboratory, Richland, Washington. 237 pp.
- Piepel, G. F., and R. J. Brouns. 1982. Statistical Sampling Plans for Prior Measurement Verification and Determination of the SNM Content of Inventories. NUREG/CR-2466, U. S. Nuclear Regulatory Commission, Washington, D.C.
- Raj, D. 1972. The Design of Sample Surveys. McGraw-Hill Book Co., New York. xii + 390 pp.
- Ripley, B. D. 1981. Spatial Statistics. John Wiley and Sons, New York. ix + 252 pp.
- Rohde, C. A. 1976. "Composite Sampling." Biometrics 32:273-282.
- Schaeffer, D., H. W. Kerster and K. G. Janardan. 1982. "Monitoring Toxics by Group Testing." Environ. Manage. 6(6):467-469.
- Scheffe, H. 1959. The Analysis of Variance. John Wiley and Sons, New York. xvi + 477 pp.
- Sedransk, J. 1965a. "A Double Sampling Scheme for Analytical Surveys." J. Amer. Stat. Assoc. 60:985-1004.
- Sedransk, J. 1965b. "Analytical Surveys With Cluster Sampling." J. Royal Stat. Soc. B 27(2):264-278.
- Sedransk, J. 1966. "An Application of Sequential Sampling to Analytical Surveys." Biometrika 53(1/2):85-97.
- Siefkin, D., G. Pangburn, R. Pennifill and R. J. Starmer. 1982. Site Suitability, Selection and Characterization. Branch Technical Position -- Low Level Waste Licensing Branch. NUREG/CR-0902, U. S. Nuclear Regulatory Commission, Washington, D.C.
- Slonim, M. J. 1960. Sampling. Simon and Schuster, New York. xiii + 144 pp.
- Smith, R. D., and E. F. Hawkins. 1980. "Low-Level Radioactive Waste Modeling Needs From the USNRC Perspective." In Modeling and Low-Level Waste Management: An Interagency Workshop, pp. 31-39. CONF-80127, NTIS, Springfield, Virginia.

- Sophocleous, M., J. E. Paschetto and R. A. Olea. 1982. "Ground-Water Net-work Design for Northwest Kansas, Using the Theory of Regionalized Variables." Groundwater 20(1):48-58.
- Steel, R. G. D., and J. H. Torrie. 1980. Principles and Procedures of Statistics, A Biometrical Approach. 2nd ed. McGraw-Hill Book Co., New York. xxi + 633 pp.
- Stoline, M. R. 1981. "The Status of Multiple Comparisons: Simultaneous Estimation of All Pairwise Comparisons in One-Way ANOVA Designs." Amer. Statist. 35:134-141.
- Stuart, A. 1976. Basic Ideas of Scientific Sampling. 2nd ed. C. Griffin and Co., London. 106 pp.
- Sudman, S. 1976. Applied Sampling. Academic Press, New York. x + 249 pp.
- Sukhatme, P. V., and B. V. Sukhatme. 1970. Sampling Theory of Surveys with Applications. 2nd ed. Iowa State University Press, Ames, Iowa. xvi + 452 pp.
- Thomas, J. M., J. A. Mahaffey, K. L. Gore and D. G. Watson. 1978. "Statistical Methods Used to Assess Biological Impact at Nuclear Power Plants." J. Environ. Manage. 7:269-290.
- Tukey, J. W. 1953. The Problem of Multiple Comparisons. Princeton University, Princeton, New Jersey.
- Urquhart, N. S. and D. S. Weeks. 1978. "Linear Models in Messy Data: Some Problems and Alternatives." Biometrics 34:696-705.
- Volpi, G., and G. Gambolati. 1978. "On the Use of a Main Trend for the Kriging Technique in Hydrology." Adv. Water Res. 1(6):345-349.
- Watson, G. S. 1971. "Trend-Surface Analysis." J. Mathematical Geol. 3:215-226.
- White, G. C. 1978. Estimation of Expected Value and Coefficient of Variation for Lognormal and Gamma Distributions. LA-7393-MS, Los Alamos Scientific Laboratory, Los Alamos, New Mexico. 82 pp.
- White, G. C., and T. E. Hakonson. 1979. "Statistical Considerations and Survey of Plutonium Concentration Variability in Some Terrestrial Ecosystem Components." J. Environ. Qual. 8(2):176-182.
- White, G. C., J. C. Simpson and K. V. Bostick. 1980. Studies of Long-Term Ecological Effects of Exposure to Uranium V. AFATL-TR-79-101, Air Force Armament Laboratory, Eglin Air Force Base, Florida. 20 pp.

- White, G. C. 1981. Kriging Analysis of Uranium Concentrations in Test Area C-74L, Eglin Air Force Base, Florida. AFATL-80-135, Air Force Armament Laboratory, Eglin Air Force Base, Florida. 8 pp.
- Williams, W. H. 1978. A Sampler on Sampling. John Wiley and Sons, New York. xv + 254 pp.
- Yates, F. 1981. Sampling Methods for Censuses and Surveys. 4th ed. MacMillan Publishing Co., New York. xvi + 458 pp.

NRC FORM 335 <small>(11-81)</small>		U.S. NUCLEAR REGULATORY COMMISSION BIBLIOGRAPHIC DATA SHEET		1. REPORT NUMBER (Assigned by DDC) NUREG/CR-4162 PNL-4804	
4. TITLE AND SUBTITLE (Add Volume No., if appropriate) Survey of Statistical and Sampling Needs for Environmental Monitoring of Commercial Low-Level Radioactive Waste Disposal Facilities				2. (Leave blank)	
7. AUTHOR(S) L. L. Eberhardt J. M. Thomas				5. DATE REPORT COMPLETED MONTH: March YEAR: 1985	
9. PERFORMING ORGANIZATION NAME AND MAILING ADDRESS (Include Zip Code) Pacific Northwest Laboratory P.O. Box 999 Richland, Washington 99352				DATE REPORT ISSUED MONTH: July YEAR: 1986	
12. SPONSORING ORGANIZATION NAME AND MAILING ADDRESS (Include Zip Code) Division of Radiation Programs and Earth Sciences Office of Nuclear Regulatory Research U.S. Nuclear Regulatory Commission Washington, D.C. 20555				6. (Leave blank)	
				8. (Leave blank)	
				10. PROJECT/TASK/WORK UNIT NO.	
				11. FIN NO. B2461	
13. TYPE OF REPORT Technical		PERIOD COVERED (Inclusive dates) November 1982 - October 1984			
15. SUPPLEMENTARY NOTES				14. (Leave blank)	
16. ABSTRACT (200 words or less) <p>This project was designed to develop guidance for implementing 10 CFR Part 61 and to determine the overall needs for sampling and statistical work in characterizing, surveying, monitoring, and closing commercial low-level waste sites. When cost-effectiveness and statistical reliability are of prime importance, then double sampling, compositing, and stratification (with optimal allocation) are identified as key issues. If the principal concern is avoiding questionable statistical practice, then the applicability of kriging (for assessing spatial pattern), methods for routine monitoring, and use of standard textbook formulae in reporting monitoring results should be reevaluated. Other important issues identified include sampling for estimating model parameters and the use of data from left-censored (less than detectable limits) distributions.</p>					
17. KEY WORDS AND DOCUMENT ANALYSIS Statistics Kriging Sampling for Pattern Monitoring Compositing Estimating Model Parameters Systematic Sampling			17a. DESCRIPTORS		
17b. IDENTIFIERS/OPEN-ENDED TERMS					
18. AVAILABILITY STATEMENT Unlimited		19. SECURITY CLASS (This report) Unclassified		21. NO. OF PAGES	
		20. SECURITY CLASS (This page) Unclassified		22. PRICE \$	