A FRAMEWORK FOR DEVELOPMENT OF
DATA ANALYSIS PROTOCOLS
FOR GROUNDWATER QUALITY MONITORING SYSTEMS

Submitted by

Nadine C. Adkins

Department of Agricultural and Chemical Engineering

In partial fulfillment of the requirements

for the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 1992

COLORADO STATE UNIVERSITY

October 16, 1992

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY NADINE C. ADKINS ENTITLED A FRAMEWORK FOR DEVELOPMENT OF DATA ANALYSIS PROTOCOLS FOR GROUNDWATER QUALITY MONITORING SYSTEMS BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Jim C. Loftis

Hariharan Iyer

Harry F. Bell

Robert C. Ward
Advisor

Jerry M. Lenz
Department Head

ABSTRACT

A FRAMEWORK FOR DEVELOPMENT OF DATA ANALYSIS PROTOCOLS FOR GROUNDWATER QUALITY MONITORING SYSTEMS

Protocols for field sampling and laboratory sampling are used on a routine basis to produce accurate and precise water quality data. Efforts are now being focused on providing decision makers with the information they need from that data. Statistics is one method of extracting information from data. There are no widely accepted protocols for statistically analyzing groundwater quality data. Due to the wide variety of field conditions encountered in groundwater monitoring, a general protocol would be of limited use. What is needed is a set of guidelines for writing site specific data analysis protocols.

A framework for developing data analysis protocols (DAPs) is presented in this thesis. The framework is essentially a "how-to" manual for protocol writers. It is designed to be concise, easy to use, and based on the current state-of-the-art. The focus of the framework is the analysis of groundwater quality data at hazardous waste facilities.

Detailed background information is presented for the framework. The four main issues that are addressed include: information goals, data record attributes, and choice and interpretation of statistical results. There is a great deal of confusion in the water quality community regarding these issues. This thesis does not attempt to resolve that confusion. Instead, the goal was to sort out the areas of conflict and uncertainty, and present them in a clear manner. Recommendations are provided where possible.

The framework was used to write a data analysis protocol for an IBM semiconductor manufacturing plant in Hopewell Junction, New York. The combination of flexibility in the basic framework and the availability of detailed background information was quite effective. It allowed the data analysis protocol to be site specific and scientifically defensible.

Nadine C. Adkins
Department of Agricultural and Chemical Engineering
Colorado State University
Fort Collins, CO 80523
Fall 1992

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank a few of the many individuals who have contributed to this thesis. First, I am fortunate to have a very talented academic committee. Harry Bell provided me with "real world" insights of groundwater remediation and environmental compliance that would be impossible to acquire from books or journals. Harry was a major contributor to the case study.

Hari Iyer taught me classical statistics and then showed me why it doesn't always work for analyzing water quality data. Hari has a genuine interest in the practical application of statistics.

Jim Loftis contributed to my thesis primarily through his knowledge of statistical analysis of water quality data. His current research on serial correlation, trend analysis, time scale and multivariate analysis was particularly useful.

Robert Ward provided consistent guidance and inspiring conversation. The success of my thesis can be largely attributed to his assistance. Robert's experience with water quality management and statistical analysis was invaluable.

The following individuals supplied key information: Graham McBride of the Water Quality Centre in New Zealand; John Harcum, Jane Harris and Bob Montgomery, former CSU students; Dennis Lettenmaier with the University of Washington and Ed Gilroy, Dennis Helsel and Robert Hirsch with the USGS. Sharon Patterson spent many hours reviewing the formatting aspects of my thesis.

IBM Corporation provided generous financial assistance for which I am quite grateful.

On a more personal level, I would like to thank my friends and family. My father put aside his rather old-fashioned ideas about women and decided that my place was in the office (or the field) as an engineer. My mother encouraged me to apply for the IBM fellowship and was always available for pep talks during my time at CSU.

My brother Butch gave me a collection of his paintings and photographs. Their stunning portrayal of nature is a reminder of the importance of pursuing an environmental career. My sister Cecelia was understanding of the pressures involved in graduate school. She sent me a personalized coffee mug with the caption: "It's just one d--deadline after another."

I would like to extend special thanks to my partner, friend and fellow student, Dale Kinnamon, who gave emotional as well as logistical support. Leslie Applegate and her Nubian goats provided a healthy source of diversion from my thesis. Thanks also to Barb, Vicki and Mary for their professional care-giving talents.

And last but not least, I want to dedicate this thesis to my dear animal friends: George the Goat, Midnight, Cosmo the Cosmic Kitten, and Lucky.

TABLE OF CONTENTS

I. INTRODUCTION

- Problem Statement 1
- Objectives 4
- Scope 5
- Organization 7

II. PROTOCOLS

- Introduction 10
- Protocols in the Literature--A Critical Review 11
- The Data Analysis Protocol Concept in Water Quality Monitoring 26
- The Need for Water Quality Data Analysis Protocols 33
- Advantages of Using Water Quality Data Analysis Protocols 34
- Characteristics of Effective Water Quality Data Analysis Protocols 36
- Conclusions 37

III. IDENTIFYING INFORMATION GOALS FOR WATER QUALITY MONITORING SYSTEMS

- Introduction 39
- The Importance of Goal Identification 40
- Regulatory Information Goals 41
- Monitoring Information Goals 56

TABLE OF CONTENTS, cont.

Statistical Information Goals 59

Conclusions 60

IV. HANDLING ATTRIBUTES OF WATER QUALITY DATA RECORDS

Introduction 62

Serial Correlation 62

Seasonality 80

Censoring 92

Nonnormality 109

Data Record Attributes that Result from Difficulties in the Monitoring Program 116

Conclusions 120

V. CHOOSING AND INTERPRETING STATISTICAL METHODS FOR ANALYZING WATER QUALITY DATA

Choosing Statistical Analysis Methods 122

Graphical Methods 131

Estimation 146

Hypothesis Testing 154

Conclusions 168

VI. A FRAMEWORK FOR DEVELOPMENT OF GROUNDWATER QUALITY DATA ANALYSIS PROTOCOLS

Introduction 169

Define the Problem 169

Gain Support for the Protocol 170

Write the Introduction 170

Include Procedures for Data Handling 171

TABLE OF CONTENTS, cont.

State Information Goals 171

Specify Procedures for Handling Data Record Attributes 171

Choose Data Analysis Methods 173

Describe How Statistical Results Should be Interpreted 173

Specify Reporting Procedures 174

Discuss How the Results of Statistical Analysis Should be Utilized in Decision Making 175

Prepare a Summary Sheet for Each Information Goal 175

State Procedures for Reviewing the Protocol 176

Include Items to Clarify the Data Analysis Protocol 176

Summary 176

VII. APPLICATION OF THE DESIGN FRAMEWORK--A CASE STUDY

Introduction 178

Background 179

Discussion 184

VIII. SUMMARY AND CONCLUSIONS

Summary 189

Conclusions 190

Recommendations for Further Work 191

Concluding Remarks 192

REFERENCES 193

TABLE OF CONTENTS, cont.

APPENDIX I:	Initial submittal of groundwater monitoring proposal.	APP-1
APPENDIX II:	Negotiations between NYSDEC and IBM regarding modifications to the intial groundwater monitoring proposal.	APP-21
APPENDIX III:	Final submittal of groundwater monitoring proposal.	APP-28
APPENDIX IV:	Data analysis protocol for groundwater monitoring conducted in response to NYS 373 permit--IBM East Fishkill Facility.	APP-30

LIST OF TABLES

IV-1	Methods for dealing with seasonality (Ward *et al.*, 1988).	85
IV-2	Limits for describing statistical properties of low-level data.	97
IV-3	Definitions of limits for describing statistical properties of low-level data.	98
IV-4	Recommended methods for statistical analysis of censored data. Adapted from Helsel and Hirsch (1992).	107
V-1	Estimators of distributional properties.	147
V-2	Estimators of difference in location based on two samples.	148
V-3	Estimators of linear dependence and monotonic trend.	149
V-4	Tests for goodness of fit.	156
V-5	Tests for location--two samples.	157
V-6	Tests for location--more than two samples.	158
V-7	Tests for linear dependence and monotonic trend.	159
V-8	Tests for correlation, equal variance and serial correlation.	160
V-9	Decisions in hypothesis testing. Adapted from Iman and Conover (1983).	166

LIST OF FIGURES

I-1	The role of statistical analysis in relation to modeling and risk assessment.	8
IV-1	Graphic depiction of the relationship between the criterion of detection (COD), α and the PDF of H_o.	100
IV-2	Graphic depiction of the relationship between the criterion of detection (COD), α and β when COD = LOD.	101
IV-3	Graphic depiction of the relationship between the criterion of detection (COD), limit of detection (LOD), α and β when $\alpha = \beta$.	102
V-1	Frequency histogram.	134
V-2	Hanging histogram.	135
V-3	Stem and leaf diagram.	136
V-4	Boxplot.	137
V-5	Normal probability plot.	138
VI-1	Main components of the framework for development of data analysis protocols for groundwater quality monitoring systems.	177

CHAPTER I

INTRODUCTION

PROBLEM STATEMENT

In response to rising environmental concern, the number of groundwater quality monitoring systems in the U.S. is steadily increasing. These monitoring systems are established to meet regulatory as well as internal management needs, but rarely are those needs articulated in terms of the nature of information to be obtained or how the information is to be used in decision making. Consequently, large quantities of high quality data are being collected without a clear definition of program goals, data analysis procedures, reporting formats or types of decisions to be made.

Groundwater quality managers are becoming increasingly aware of the need to develop documented strategies (i.e., protocols) for statistically analyzing data obtained from monitoring programs. Data analysis protocols (DAPs) help to ensure that data obtained from monitoring programs can be translated into useful information that meets program goals. Assessing the effectiveness of remediation efforts and detecting the presence of contaminants are examples of program goals.

Groundwater quality regulations often contain language similar to the following from the New York Code of Rules and Regulations:

> "In conjunction with a corrective action program, the owner or operator must establish and implement a ground water monitoring program to demonstrate the effectiveness of the correction action program" (Sec. 373-2.6.k.4 of 6NYCRR).

Such regulatory directives usually do not include details of how to analyze groundwater quality data to assess the effectiveness of remediation efforts. It is left to the permittee to decide how to analyze the data. Although a list of proposed statistical methods is generally included in the overall remediation plan submitted by the permittee, it is not standard practice to develop and implement data analysis protocols. As a result, water quality managers often find themselves spending large quantities of money on monitoring programs that do not yield the information they need for decision-making.

Ward *et al.* (1990) have defined water quality monitoring systems in terms of the flow of information through the system. The flow of information begins with sample collection and ends with information utilization as shown below:

1. Sample Collection
2. Laboratory Analysis
3. Data Handling
4. Data Analysis
5. Reporting
6. Information Utilization

The first three components deal primarily with data collection whereas the last three deal with information generation. Researchers have traditionally focused their efforts on data collection issues. Research on data collection has been used to develop standard procedures for sample collection, laboratory analysis and data handling. The use of these standard procedures has resulted in the collection of groundwater quality data that has a high degree of accuracy. Accurate data is, of course, a necessary prerequisite to the generation of useful information.

Water quality researchers and professionals are beginning to switch their focus away from data collection and towards information generation. Some research on information generation activities has already been conducted, particularly in the area of statistical analysis. The research is difficult to apply to monitoring system design, however, because it is so widely scattered. A few attempts have been made to produce standardized sets of procedures for information generation known as "data analysis protocols." (The term is somewhat of a misnomer, because it refers to more than just data analysis itself. It also encompasses identification of information goals, handling of data record attributes, interpretation of statistical results and reporting.) These data analysis protocols (DAPs), however, tend to be incomplete and/or too general to be useful to monitoring system designers.

Due to the wide variety of information needs and site conditions, it is impractical to expect a single DAP to be suitable for all groundwater quality monitoring systems. What is really needed is a framework that can be used to develop DAPs that are program specific. No generally acceptable design framework for the development of groundwater quality data analysis protocols exists today.

OBJECTIVES

The primary objective of this thesis is to present a framework for the development of groundwater quality data analysis protocols. The protocols are intended to be program specific and should be written during the initial phases of monitoring system design. Application of the data analysis protocols should result in the generation of information that is employed in decision making.

Four main components of the framework that are expanded upon in this thesis are: (1) identification of information goals, (2) handling of data record attributes, (3) choice of statistical analysis methods and (4) interpretation of statistical results.

The second objective of this research is to explore the concept of protocols by conducting an extensive literature review. A third thesis objective is to demonstrate the practical application of the DAP design framework by conducting a case study. A case study was conducted for the IBM semiconductor manufacturing plant in East Fishkill, New York.

SCOPE

The intended users of the framework are individuals who will write groundwater quality DAPs for hazardous waste facilities. Such individuals will probably work directly for the facility or be hired as consultants. The framework can be used to develop DAPs for internal use and/or for submittal to regulators.

Economic factors are not specifically discussed. It is reasonable to assume, however, that the use of data analysis protocols will produce economic benefits.

The framework should only be used to develop DAPs for chemical groundwater quality monitoring programs. Surface water quality and biological groundwater quality were not considered in the formulation of the framework.

Although the protocol design framework emphasizes the use of statistical methods of data analysis, there are situations where alternative methods may be more appropriate. Examples of other data analysis methods are presented in the case study (Chapter VII).

This thesis does not describe the many complexities that contribute to the making of water quality legislation and resulting regulations. Instead, it emphasizes how regulations should be dealt with once they are promulgated. In particular, it emphasizes meeting groundwater monitoring and data analysis requirements being placed on industrial sites.

DAPs developed from the framework should be written prior to data collection. Existing data, however, may be used for data characterization or to confirm analytes.

It is assumed in this thesis that only high quality data is obtained from the laboratory. The production of high quality data has been a priority for several years in the field of water quality. The use of sampling protocols and laboratory analytical protocols (e.g., ASTM procedures) has become routine. The assumption of high quality data should be reasonable for hazardous waste sites that have an effective QA/QC program.

The focus of this thesis is on practical application of statistical methods. It is aimed more towards management of water quality than the theory of statistics.

The framework is intended for development of DAPs that produce information on **what** the current state of water quality is. The framework is not designed to produce protocols that address the question of **why** water quality is the way it is.

There is a great deal of confusion and uncertainty regarding the use of statistics in the analysis of water quality data. To cope with that situation, this thesis deals with the current state-of-the-art and points out areas that are currently under debate.

A basic knowledge of statistics is assumed. If a reader is not familiar with statistics, however, they should still be able understand the thesis by referring to an introductory text and/or references noted to contain explanatory detail.

This thesis addresses the question of how to use statistical analysis to obtain information from data. It does not address the use of modeling or risk assessment. Figure I-1 shows the role of statistical analysis in relation to modeling and risk assessment.

ORGANIZATION

Protocols are discussed in Chapter II. Chapters III through V cover the main aspects of the DAP framework: identification of information goals, handling of data record attributes, and choice and interpretation of statistical analysis methods. The DAP framework is presented in Chapter VI and the case study is discussed in Chapter VII.

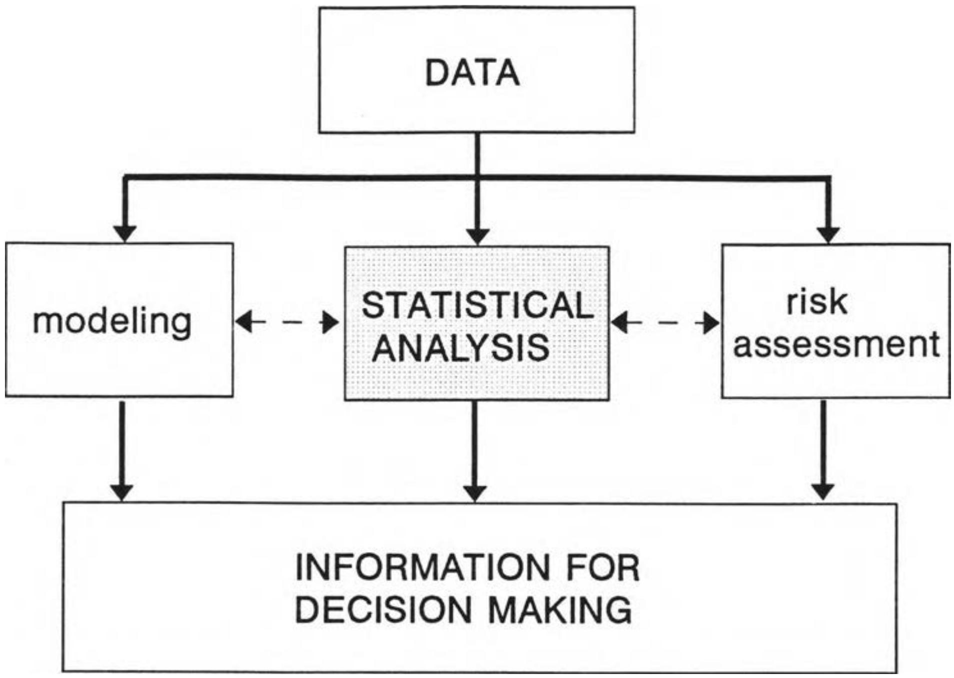


FIGURE I-1. The role of statistical analysis in relation to modeling and risk assessment.

Conclusions and recommendations for further work are given in the final Chapter. The case study is presented in Appendices I-IV.

CHAPTER II

PROTOCOLS

INTRODUCTION

Definition

The term "protocol" is used in several ways. Many people think of a protocol in terms of diplomacy, either as a signed agreement between negotiating parties or as the proper rules of etiquette used by high officials. The accepted definition in scientific literature, though, is simply: **a standardized set of procedures for doing something.** For example, in hydrology, there are protocols for constructing unit hydrographs and for designing culverts. The American Society for Testing and Materials (ASTM) publishes hundreds of protocols for laboratory and field testing of various substances including water and soil.

Organization

An evaluation of what can be learned from existing protocols is presented in the first part of this chapter, followed by a review of the data analysis protocol concept in water quality monitoring. Then the need for groundwater quality DAPs is discussed. Finally, the advantages and characteristics of a successful groundwater quality DAPs are identified.

PROTOCOLS IN THE LITERATURE--A CRITICAL REVIEW

Objective

A literature review of protocols from a variety of disciplines was conducted in order to gain a better understanding of how to develop water quality data analysis protocols (DAPs). The review was not intended to be an exhaustive or precise study, but rather a means to get a general feel for what makes a protocol successful.

Research Strategy

References for the majority of articles included in this literature review were obtained from CDROM catalogs in the CSU research library (using "protocol" or "protocol development" as key words). References were acquired for a large variety of disciplines including microbiology, medicine, toxicology, zoology, environmental engineering, chemistry and anthropology to name a few.

A few documents were reviewed that presented protocols but did not identify them as such. *The Consumer Price Index: History and Techniques* (U.S. Dept. of Labor, 1967) and *Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities* (U.S. EPA, 1989) are examples.

Results

This broad-based literature review yielded a surprisingly large amount of information that can be used to develop groundwater quality data analysis protocols (DAPs). Many disciplines, such as medicine and biology, have been in existence longer than groundwater engineering, and consequently, are further along in the standardization processes that generally lead to formulation of protocols.

Observations from the literature review are presented below in a question and answer format. The applicability of these observations to the development and implementation of groundwater quality DAPs is discussed.

QUESTION: What can one do to ensure the success of a protocol?

ANSWERS:

1. **Gain an understanding of the problem:** In an article on bioassay protocols (Davis, 1977), the author states "The obvious precursor to standardization is a thorough knowledge of the various chemical, physical and biological factors that affect bioassay results..." This concept is applicable to writing a groundwater quality DAP because statisticians are very adamant that the problem must be understood before statistical analysis is conducted (Zahn and Isenberg, 1983).

Conducting a literature review and doing a pilot study are two ways to increase understanding.

2. Hold a workshop: Workshops are a commonly used method for getting people together from a variety of disciplines to discuss complex issues. Workshops have been held specifically to discuss protocol standardization for Puget Sound, Washington (Armstrong and Becker, 1986), development of a scientific protocol for ocean dumpsite designation (Reed and Bierman, 1983) and protocol development for potable water reuse criteria (Cotruvo *et al.*, 1982). Such an approach is appropriate for writing protocols that will be used on a regional or national basis. A regular meeting as opposed to a formal workshop would probably be more practical for discussing a groundwater quality DAP for a specific hazardous waste site.

3. Circulate a draft: A few authors mentioned that they passed around a draft for comments before writing the final protocol (Armstrong and Becker, 1986; Reed and Bierman, 1983 and Schuk, 1986). Circulating a draft should be appropriate for just about any type of protocol. For one thing, the chances of the protocol being accepted are increased if users and managers are involved in the development process. Also, a protocol by its nature is a consensus agreement on how something should be done.

4. Conduct a literature review: Literature reviews were conducted for a nutritional risk screening protocol (Hedberg *et al.*, 1988), a hazardous waste reactivity testing protocol (Wolbach *et al.*, 1984) and a protocol for measuring hydrolysis rate constants in aqueous solutions (Ellington *et al.*, 1988). A couple of reasons to conduct a literature review are: (1) to gain a better understanding of the problem and (2) to identify commonly used methods that can be referred to in the protocol. For these reasons, a literature review is an appropriate activity to undertake prior to writing a groundwater quality DAP. Also, statistical analysis of groundwater quality data is a complex and relatively new field that is not adequately covered in any one document.

5. Conduct pilot studies: Conducting a pilot study is one way of acquiring a better understanding of the problem before the final protocol is written (Duke and Merrill, 1981 and Schroder and Taylor, 1980). In water quality monitoring, a pilot study would be helpful for data characterization, which in turn would be useful in choosing appropriate statistical methods. Because groundwater quality monitoring is so expensive, however, a pilot study conducted exclusively for data characterization may not be economically feasible. The options would be to deal with characterization issues during the actual monitoring program or to examine data from past programs.

QUESTION: How focused are successful protocols?

ANSWER:

Most protocols encountered in the literature review are narrowly focused. Some examples are listed here:

- A hospital radiation emergency protocol for Victoria General Hospital (Aldrich, 1988).
- A pronghorn hand-rearing protocol for the Los Angeles Zoo (Brinkley, 1987).
- An analytical protocol for the multimedia characterization of polychlorinated dibenzodioxins and dibenzofurans by high-resolution gas chromatography/high-resolution mass spectrometry (Tondeur *et al.*, 1989).

Because these protocols are narrowly focused, it is possible for the authors to be quite specific. A broadly focused protocol such as one for aquifer decision making (Canter and Knox, 1986) is necessarily general. In fact, the author states "The procedure is not intended to be a set of explicit instructions, but rather a general approach which, when modified, could be applied to a wide variety of ground water pollution problems."

A protocol for groundwater quality data analysis would be the most effective if it was narrowly focused. Hydrogeologic conditions are so variable that a protocol that was developed for a particular site would most likely be inappropriate at a different site. Even at the same site, a protocol written for a specific monitoring program may not work for the next monitoring program due to changing objectives.

QUESTION: What features enhance the clarity and usefulness of a protocol?

ANSWERS:

1. **Flow charts:** Flow charts are often used to visually convey the structure of all or portions of a protocol (Reed and Bierman, 1983; Thomas, 1987 and U.S. EPA, 1989). A reader can glance at a well designed flowchart and quickly comprehend the basic organization of a process or procedure. An example of simple yet informative flow charts can be seen in a protocol that was developed for ocean disposal site designation (Reed and Bierman, 1983).

2. **Glossaries:** An EPA handbook titled *Protocols for Short Term Toxicity Screening of Hazardous Waste Sites* (Greene *et al.*, 1988), begins with a glossary. There is enough confusion over statistical and water quality terminology that a glossary would surely be a welcome addition to a groundwater quality DAP.

3. **Sample forms:** An excellent example of the inclusion of sample forms in a protocol is the EPA document, *Interim Protocol for Measuring Hydrolysis Rate Constants in Aqueous Solutions* (Ellington *et al.*, 1988). The forms, which are presented as appendices, are for generation of data and for calculation of rate constants and activation energies. Inclusion of sample forms appears to be a simple and effective approach to ensure that protocol users generate data and perform calculations in a consistent manner. A similar approach may enhance the effectiveness of groundwater quality DAPs.

QUESTION: What are some recommendations for development of protocols?

ANSWERS:

1. **Clearly define program goals:** The following statement is made in an article that presents a framework for the development of protocols to monitor

energy systems in buildings: "Perhaps the most critical activity to the success of a monitoring project is correctly and accurately stating its goals, objectives, and the specific research questions to be answered" (Misuriello, 1987). There is widespread agreement among water quality professionals that it is important to clearly define the goals of groundwater monitoring programs. Statisticians also agree on the importance of well defined goals (Hunter, 1981). It would seem logical, therefore, that a protocol for the statistical analysis of groundwater quality data should include a carefully thought out statement of program objectives.

2. **Use accepted methods:** Specification of accepted (i.e., proven and widely used) methods in protocols is promoted in several of the articles that were reviewed. For example, an article that presents a protocol for the identification of toxic fractions in industrial wastewater influents, includes the following statement: "To ensure practicality of the protocol as a routine procedure, the fractionation was based on simple and widely used laboratory techniques..." (Gasith *et al.*, 1988). A similar statement is made in an article on the development of bioassay protocols: "Finally, each test incorporated into the protocol was to be

acceptable by the scientific community as valid, reliable and accurate with a sufficiently large existing data base to facilitate the interpretation of results."

Accepted statistical procedures should be specified in a groundwater quality DAP to promote routine application of the protocol and to facilitate the interpretation of results. It should be remembered, however, that the use of statistics in analyzing water quality data is a rapidly advancing field. Methods that are "accepted" one year may be replaced by better methods the next year. A protocol author should keep informed of the latest advances and revise the protocol accordingly.

3. Allow for revision: A large number of articles that were reviewed included statements indicating that the protocols should be revised as further knowledge is gained. For example, in reference to a protocol for testing microbiological water purifiers: "(it) is intended to be a living document, subject to revision and update as new knowledge and technology arise" (Schaub and Gerba, 1988). When discussing a protocol for determining lake acidification pathways, the author states "...it provides a general framework which can be challenged and expanded upon through the addition of other evidence..." (Marmorek *et al.* 1989). A protocol for biological testing was revised and refined to solve problems revealed by pilot studies and implementation (Duke and Merrill, 1981). In reference to a pronghorn hand-rearing protocol: "Over the years we have made modifications to the original protocol as our needs and experience demanded" (Brinkley, 1987).

Allowing for revision of groundwater quality DAPs is advisable. As mentioned previously, the use of statistics to analyze groundwater quality data is a rapidly evolving field. Also, as more data is obtained in a monitoring program, it may become evident that a change in statistical methodology is appropriate. A systematic revision mechanism should be incorporated directly into the protocol to prevent haphazard and unauthorized changes.

4. Acknowledge limitations: Some of the articles that were reviewed acknowledged limitations of the protocols. For example, in reference to a protocol for determining lake acidification pathways: "The protocol cannot, however, distinguish between low pH lakes which originally contained little DOC, and those which lost it" and "The decision protocol presented in this paper is still very preliminary" (Marmorek *et al.*, 1989). The author of a nutritional risk screening protocol states, "Although the screening program has proved to be an effective tool to identify patients at nutritional risk, it does not substitute for the dietician's clinical subjective experience" (Hedberg, *et al.*, 1989).

It is important to acknowledge the limitations of a groundwater quality DAP and the associated statistical procedures so that managers can take them into consideration when making decisions. People tend to place more confidence in statistical results than what is warranted.

QUESTION: What are some advantages of using protocols?

ANSWERS:

1. **They save money:** Many protocols are designed to save money. For example, when faced with budget cuts, the EPA developed a protocol to more effectively interpret data from EIS documents to determine possible effects or hazards associated with ocean disposal operations (Reed and Bierman, 1983).

A groundwater quality DAP can save money in several ways. For one thing, data is not collected unless it is needed for a specific purpose. Also, the maximum amount of information is obtained from the data that is collected.

2. **They save time:** The need to conduct a project in a timely manner is a common motivator for protocol development. Scientists at the U.S. Department of Agriculture Eastern Research Center wrote a protocol for the accumulation of fatty acid data from multiple tissue samples. The protocol was developed to "allow for an unbiased handling of the samples, proper record-keeping, and - most importantly - the timely completion of an otherwise unmanageable task" (Maxwell and Marmer, 1983).

A protocol for analyzing groundwater quality data results in time-savings because data is not collected that is not needed. Furthermore, time is not wasted "reinventing the wheel" whenever data needs to be analyzed.

3. **They produce comparable results:** A frequently cited advantage of protocols is that data collected by different investigators can be compared. At the start of a regional effort to protect and manage the environment of Puget Sound, it became apparent that

data comparability would be a real problem because so many different approaches were being used to sample and analyze the same parameter (Becker and Armstrong, 1988). A series of protocols were written to alleviate this problem. As stated by the protocol authors, "Perhaps the best way to ensure that data collected during different studies are comparable is to encourage all investigators to use standardized sampling and analysis protocols whenever possible."

The advantage of comparable results is probably the most evident when monitoring is conducted by many different people on a regional basis. Comparable results are still an advantage, however, even when a groundwater quality DAP is written for a specific monitoring program at a single site. Because a protocol documents how statistical analysis results are obtained, future investigators are able to determine if results can be compared.

4. Reliable data is obtained: Several protocols are commended by their authors because they produce reliable data. Some examples are: (1) a protocol for measuring hydrolysis rate constants in aqueous solutions (Ellington *et al.*, 1988), (2) a protocol for measuring microbial transformation rate constants for suspended bacterial populations in aquatic systems (Steen, 1988) and (3) a protocol for collection of saliva samples under field conditions (Lipson and Ellison, 1987).

The chances of reliable information being obtained from a groundwater quality monitoring program are increased if a DAP is used. The process of writing a DAP forces the writer to carefully consider which statistical methods to choose and how to interpret the results. This is in contrast to the haphazard, last-minute application of statistics that often occurs in water quality monitoring. DAPs also lead to more reliable information because they allow data analysis procedures to be audited.

5. **Results are reproducible:** A few authors cited reproducibility of results as an advantage of their protocol. For example, a treadmill protocol for measurement of aerobic parameters (Cowell, 1989) and a protocol for testing effects of toxic substances on plants (Thompson *et al.*, 1981) both generate reproducible results.

The generation of reproducible results is a significant advantage of groundwater quality DAPs. Too often statistical conclusions are reported in such a way that the reader has no idea of how they were obtained.

Summary

Concepts from protocols in other disciplines can be used to develop protocols for analyzing groundwater quality data. These concepts are summarized here.

Prior to writing a final protocol, it is important to understand the issue that is being addressed. For example, radiation emergencies and hospital policy should be understood before writing a hospital radiation emergency protocol. Conducting a literature review or a pilot study are two ways to improve understanding. It is also beneficial to gather ideas from other people before writing the final protocol. Input from others can be obtained by holding a workshop and/or circulating a draft.

The majority of protocols presented in the literature are narrowly focused. For example, a protocol has been developed for radiation emergencies at a specific hospital in Rhode Island.

Flow charts, glossaries and sample forms are three features that are used to enhance the clarity and usefulness of protocols. Recommendations for writing a protocol include: (1) state objectives, (2) use accepted methods, (3) allow for revision and (4) acknowledge limitations.

Certain advantages of protocols are frequently mentioned in the literature. Protocols save time and money. Also, they produce comparable and reproducible results, as well as reliable data.

THE DATA ANALYSIS PROTOCOL CONCEPT IN WATER QUALITY MONITORING

Protocols that deal with monitoring system design, sampling and laboratory analysis are routinely used in the field of water quality monitoring. Protocols to statistically analyze water quality data, however, are still in the early stages of development.

Five documents that attempt to standardize various facets of water quality data analysis are reviewed here. The purpose of the review is to gain knowledge that may be helpful in developing DAPs for groundwater quality monitoring programs at hazardous waste sites. Specific details of statistical methods are discussed in subsequent chapters. The five documents that are reviewed are:

1. Nonparametric Tests for Trend Detection in Water Quality Time Series (Berryman *et al.*, 1988)
2. Methodology to Derive Water-Quality Trends for Use by the National Water Summary Program of the U.S. Geological Survey (Lanfear and Alexander, 1990)
3. Water-Quality Data Analysis Protocol Development (Harcum, 1990)

4. Groundwater Quality: A Data Analysis Protocol (Ward *et al.*, 1988)
5. Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities: Interim Final Guidance (U.S. EPA, 1989)

Nonparametric Tests for Trend Detection in Water Quality Time Series (Berryman *et al.*, 1988)

The main objective of this paper is to present a method for systematically choosing the appropriate nonparametric statistical test for analyzing a particular time series. Tests are selected based on: (1) the dependence and sources of dependence (trend, periodicity, or persistence on random fluctuations) in the time series and (2) whether the trend is monotonic or step.

The protocol for selecting tests for monotonic trends is presented in the form of a flowchart as well as verbally. The procedure can be modified for use with step trends.

The authors stress that the test selection procedure is a general guide that should not be "followed blindly" because user judgement is sometimes required. They point out that the role of user judgement will be diminished in the future as increased knowledge allows for better definition of test selection criteria.

In addition to presenting the test selection protocol, the authors discuss related issues such as power and sample size. They briefly review methods to model time series, as well as graphic and parametric methods to detect trends. Results of a case study using the test selection protocol to choose tests for detecting lake acidification is presented.

Methodology to Derive Water-Quality Trends for Use by the National Water Summary Program of the U.S. Geological Survey (Lanfear and Alexander, 1990)

The USGS developed computer software to do the following: (1) objectively determine the suitability of a water-quality record for trend testing, (2) decide whether to do a monthly, bimonthly or quarterly trend test, (3) prepare data for analysis, (4) perform the Seasonal Kendall monotonic trend test and calculate the trend slope and (5) report the results. The software (i.e., protocol) was initially checked on data from four states. It was refined until it successfully dealt with all of the data records. After further testing and refinement, the protocol was used to conduct over 50,000 trend tests on data collected from almost 3,000 stations nationwide.

Lanfear and Alexander make this statement regarding the protocol: "Perhaps the most important lesson to be learned from the experiences of developing an automatic trend test is that software must be very `smart' if it is to cope with the myriad ways in which water-quality data are collected and recorded." Anyone who has dealt with statistical analysis of water quality data would probably agree with their observation.

Water-Quality Data Analysis Protocol Development (Harcum, 1990)

In his dissertation, Harcum suggested that a water quality data analysis protocol should consist of five components: (1) identification of information goals and transformation into water quality conditions, (2) data handling, (3) identification of data record attributes, (4) water-quality evaluation and (5) information reporting. Harcum reviewed the literature and identified those data analysis procedures that have gained wide acceptance. He proposed that a data analysis protocol could be developed simply by combining those procedures. For example, most authors agree that the Mann-Kendall tau test (or variations) should be used for monotonic trend detection, and that the Sen slope estimator should be used for monotonic trend magnitude estimation.

Harcum also proposed that a DAP could be developed by conducting simulation studies. Although he did not actually produce a protocol, Harcum conducted simulation studies that yielded important information that could be used in future protocol development efforts.

From an economic standpoint, it is unrealistic to expect monitoring system designers to conduct simulation studies for DAP development. It is more practical for them to write protocols based on information that has already been presented in the literature.

Groundwater Quality: A Data Analysis Protocol (Ward *et al.*, 1988)

A data analysis protocol is presented that contains data preparation procedures, graphical evaluation techniques and recommended statistical methods. The protocol was developed for analysis of groundwater quality data at regulated industrial facilities, but can be modified to analyze other environmental data.

The protocol has several positive features. For one thing, it is easy to use and understand. Also, it uses flowcharts and demonstrates the application of the protocol with a case study.

A disadvantage of the protocol is that it is aimed at industrial facilities in general and is therefore too broadly focused for direct use in a specific monitoring program. Also, it does not explain the logic behind many of the recommendations that are given such as which statistical methods to use.

Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities: Interim Final Guidance (U.S. EPA, 1989)

This EPA document provides guidance on the selection, use and interpretation of statistical methods to evaluate groundwater quality monitoring data at RCRA (Resource Conservation and Recovery Act) facilities. The manual is a clarification and expansion of amendments made to the RCRA regulations on October 11, 1988. Prior to those amendments, the regulations required that groundwater data be analyzed by the Cochran's Approximation to the Behrens-Fisher Student's t-test (CABF). The 1988 amendments eliminated the CABF procedure in favor of five different statistical methods. It is these five methods that are the main topic of discussion in the statistical guidance document.

In many ways, this manual represents the state-of-the-art in groundwater quality DAPs. It is the first (and only) widely accepted national effort to standardize the use of statistics in groundwater quality monitoring. Several attractive features of the document are listed here:

- The scope, purpose and intended users of the protocol are clearly defined.
- Flowcharts and example calculations are used.
- Background information is provided, such as an overview of the regulations.
- Limitations of the protocol are stated.
- A glossary of statistical terms is given.

- Frequent references for further information are provided.
- The importance of considering environmental factors such as hydrogeology and geochemistry in conjunction with statistical results is emphasized.
- The importance of understanding statistical methods prior to using them is promoted.

Even though the EPA statistical guidance document is a major improvement over previous standardization efforts, there is room for improvement. It would be more practical if it could be used either as a program-specific protocol or as a framework for developing such protocols. It is too flexible and broadly focused to be used directly for a specific monitoring program, and it is not designed to be used as a framework for protocol development. Although the document emphasizes that site-specific factors must be taken into consideration when choosing, applying and interpreting statistical methods, it does not provide guidance on how to write a protocol that incorporates those factors.

There are also some technical difficulties with the manual. It recommends parametric over nonparametric methods which is contradictory to what has been presented in the literature. (See discussion on nonparametric methods in Chapter V). It puts too much emphasis on hypothesis testing at the expense of estimation procedures and it does not adequately address the concept of statistical power (McBride, *et al.*, 1992).

THE NEED FOR WATER QUALITY DATA ANALYSIS PROTOCOLS

In the past, most research efforts in the area of water quality monitoring have been directed towards data collection rather than towards information generation. Attitudes are changing, however, and people are beginning to realize the importance of obtaining information for decision-making. This attitude is reflected in a statement made by Schubel (1987) regarding estuarine environmental monitoring, "We should spend as much on analyzing data, converting them into information and putting them in the hands of decision makers as we spend on collecting them."

Many authors have recognized the need for standardized procedures to ensure that data obtained from monitoring programs can be translated into useful information that meets program goals. For example, the following statement was made in a RCRA implementation study (U.S. EPA, 1990):

> "We must define what the RCRA program wants to measure by developing environmentally based goals and objectives. Using these milestones, we must develop an information management plan that serves as the blueprint for collecting information and for developing necessary systems... We must assemble and analyze the data so that we know where the greatest environmental

risks from hazardous waste occur, and can measure the program's success in terms of risks reduced or avoided, rather than the number of activities undertaken (permits issued, inspections performed, etc.)."

Based on the above statement, it appears that the use of data analysis protocols on RCRA projects could be quite beneficial. A DAP contains "environmentally based goals and objectives", serves as an "information management plan" and provides a means to "assemble and analyze data" to provide useful insights on the environment.

ADVANTAGES OF USING WATER QUALITY DATA ANALYSIS PROTOCOLS

The use of water quality data analysis protocols can provide the following benefits:

1. Communication between monitoring system designers, environmental managers and regulators is improved.
2. Continuity of data analysis in the face of employee turnover is facilitated.
3. Existing knowledge about water quality variables is incorporated into the monitoring program.
4. Scientific understanding plays a more important role in decision-making.
5. Monitoring system designers, environmental managers and regulators develop a better understanding of the role of statistics in obtaining water quality information.

6. The subjectivity of statistical analysis is attenuated because statistical methods are specified prior to data collection.
7. Those involved in data collection develop an understanding of the importance of producing high quality data. For example, a DAP may explain why data should not be censored in the laboratory.
8. Economic benefits are realized because only data that contributes to useful information is collected. Also, the maximum amount of information is obtained from the data that is collected.
9. Because they are well documented, data analysis procedures can be reviewed by many knowledgeable people. Too often these procedures only exist in someone's head.
10. The monitoring system design is driven by information goals rather than by politics or short term crises.
11. Data characterization is only conducted if it will contribute to the generation of useful information. For example, demonstration of normality is only necessary if the data is to be analyzed by parametric statistical methods.
12. Final decisions made during regulatory negotiations are put in writing.

13. Future investigators will know exactly how statistical results were obtained.
14. Water quality information is extracted from the data as soon as possible. It is often feasible to use a particular statistical method in the initial stages of monitoring and then switch to a more powerful method as additional data are collected. For example, time series plots may be used to "analyze" trends until enough data is available to apply formal trend tests and/or estimation procedures.
15. Reliable information is obtained because the choice of statistical methods and interpretation of results are carefully considered when the protocol is written.
16. Sampling frequencies are chosen based on information goals.

CHARACTERISTICS OF EFFECTIVE WATER QUALITY DATA ANALYSIS PROTOCOLS

Several characteristics of groundwater quality data analysis protocols that contribute to their effectiveness are identified in this section. The characteristics were chosen based on a literature review of protocols, an examination of water quality data analysis protocols, and discussions with water quality professionals and personal judgment.

1. DAPs should be narrowly focused.
2. They should include features that enhance their clarity and usefulness such as sample forms, glossaries and flowcharts.
3. Accepted statistical methods should be used.
4. Protocol limitations should be stated.
5. There should be agreement among users regarding content of the DAPs.
6. They should address the following topics:
    - Identification of information goals
    - Handling of data record attributes.
    - Graphical presentation of data.
    - Choice of data analysis methods.
    - Interpretation of results.
    - Information reporting.
    - Protocol revision.

CONCLUSION

A literature review of protocols from a variety of disciplines was conducted in order to gain a better understanding of how to develop water quality data analysis protocols (DAPs). The review produced a significant amount of useful information which is summarized in this chapter.

The data analysis protocol concept in water quality monitoring is examined by reviewing five documents that attempt to standardize various facets of water quality data analysis. The need for water quality DAPs and advantages of using them are discussed. Finally, several characteristics that contribute to effective water quality data analysis protocols are listed.

CHAPTER III

IDENTIFYING INFORMATION GOALS FOR WATER QUALITY MONITORING SYSTEMS

INTRODUCTION

Problem Statement

Information goals provide the basis for data analysis protocols. Data attribute handling, choice of statistical analysis methods, interpretation of results and reporting are all dependent on what we want to know about water quality conditions.

Identification of information goals is a three step process. First, regulatory goals are identified by meeting with regulators and reviewing regulations. Then monitoring information goals are established. Finally, if statistical methods are used to achieve the monitoring goals, specific statistical goals are developed.

The process of identifying information goals is highly dependent upon communication between DAP users. It is a process that involves asking questions, sharing ideas and developing an understanding of the problem.

Organization

The importance of goal identification is covered first, followed by a general discussion of how to identify regulatory information goals. Then, a detailed examination of information goals expressed in RCRA and CERCLA regulations is given. Finally, criteria for identifying both monitoring and statistical goals are presented.

THE IMPORTANCE OF GOAL IDENTIFICATION

The importance of information goal identification in the design of water quality monitoring systems has been noted by many authors. A few quotations regarding identification of information goals are listed here.

> "... it is simply a waste of money to monitor without a clear relationship between the information to be produced and its use within the management agency's decision making process" (Ward *et al.*, 1990).

> "A clear statement of the program's monitoring goals, objectives and environmental needs, including both narrow and broader long-term needs, is perhaps the most important section of the (environmental monitoring) strategy" (U.S. EPA, 1985).

"Many monitoring programs are ineffective because they devote too little attention to the formulation of clear goals and objectives..." (NRC, 1990).

"We must define what the RCRA program wants to measure by developing environmentally based goals and objectives" (U.S. EPA, 1990).

The development of DAPs should be driven by the information goals. Decisions on how to handle data attributes, which statistical methods to use and how to report results should be consistent with information goals.

REGULATORY INFORMATION GOALS

The Role of Conversation

Identification of information goals to meet regulatory requirements for groundwater quality monitoring systems can be a perplexing process. For one thing, environmental regulations are not written in specific enough terms to be used directly as information goals. It is left up to regulators and industry representatives to decide, on a site specific basis, what information is needed from the monitoring system and what the monitoring system can produce.

Another barrier to identifying information goals is that regulations are frequently not based on a true understanding of the problem. This deficiency is reflected in a remark made by a staff assistant in the U.S. House of Representatives: "It is a complete and utter disaster when you begin to look at the data upon which some people on Capitol Hill are talking about basing regulation - a crap shoot in many cases... We went into this issue thinking we had this area pretty well boxed in and I've personally come out so confused I don't know where to go next" (Nelson and Dowdy, 1988).

Opton (1983) explored applicant behavior as a factor in obtaining environmental permits. He found that applicants can have a great deal of control over the permit process simply by communicating effectively with regulators.

It is clear that discussions between monitoring system designers, environmental managers and regulators are essential to formulation of information goals. As stated by Wurman (1989), "Conversations are organic: their very structure is a give-and-take that allows understanding to happen."

Reviewing the Laws and Regulations

Regulations should be thoroughly reviewed prior to defining information goals. It may also be beneficial to review the applicable statutes in order to gain further insight into the intent of the regulations.

Despite the fact that laws and regulations can be vague and unscientific, it is important to review them carefully because they provide the basic "rules of the game." Although there is often plenty of room for negotiation, final water quality decisions must be made within the framework of the regulations.

Statutes are published in the U.S. Code, a multi-volume set of books that is divided into 50 titles. Each title covers a broad subject area. Most laws that pertain to groundwater monitoring, including RCRA and CERCLA, can be found in Title 42--The Public Health and Welfare.

Federal agency regulations are printed in the Code of Federal Regulations (C.F.R.), which is also a multi-volume set of books that is divided into 50 titles. EPA regulations are published in Title 40--Protection of the Environment. The C.F.R. is updated daily in the Federal Register.

Resource Conservation and Recovery Act (RCRA)

General--The Resource Conservation and Recovery Act (RCRA) regulates handling and disposal of solid waste from the point of generation to ultimate disposal (Jorgensen, 1989). RCRA is the first comprehensive piece of federal legislation that addresses the problems of hazardous waste (Hall *et al.*, 1987). Protection of groundwater from hazardous waste leachates is covered in detail by RCRA.

The first federal law that required environmentally sound solid waste disposal practices was the Solid Waste Disposal Act (SWDA) of 1965 (U.S. EPA, 1990). The SWDA was "amended" (virtually rewritten) in 1976 by the Resource Conservation and Recovery Act. The acronym "RCRA" generally refers to the 1976 Act as codified and amended (Cooke *et al.*, 1987a). Major RCRA amendments include the Used Oil Recycling Act of 1980, the Solid Waste Disposal Act Amendments of 1980 and the Hazardous and Solid Waste Amendments (HSWA) of 1984 (Cooke *et al.*, 1987a). The HSWA of 1984 substantially broadened the scope and coverage of RCRA.

Groundwater Quality Information Goals Expressed in the RCRA Statute--The RCRA statute [U.S. Code, Title 42, Sections 6901-6991(i)] mandates that EPA promulgate regulations to require groundwater quality monitoring systems at a variety of hazardous waste sites. The underlying objectives of the monitoring are alluded to in a couple of places in the statute as described here.

One of the eleven major objectives of the RCRA statute refers to the preservation of water resources:

> "The objectives of this chapter are to promote the protection of health and the environment and to conserve valuable material and energy resources by - promoting the demonstration, construction, and

application of solid waste management, resource recovery, and resource conservation systems which preserve and enhance the quality of air, water, and land resources."

EPA is authorized to require the owner of a hazardous waste site that "may present a substantial hazard to human health or the environment, (to monitor) to ascertain the nature and extent of such hazard." It is clear that EPA had very little guidance on what the information goals of groundwater quality monitoring should be.

Groundwater Quality Information Goals Expressed in the RCRA Regulations--RCRA hazardous waste regulations that apply to treatment, storage and disposal (TSD) facilities will be discussed in this paper. The TSD regulations make up the most detailed and complex category of RCRA regulations (Hall *et al.*, 1987).

There are TSD groundwater regulations for interim facilities and for permitted facilities. An interim facility is one that has not yet received a permit but has complied with certain generic performance standards that allow it to remain in operation (U.S. EPA, 1990). The owner or operator (referred to collectively as "owner") of an interim status facility is responsible for interpretation and application of groundwater regulations, whereas the owner of a permitted facility can refer to site-specific details in the permit (Hall *et al.*, 1987). Groundwater regulations and information goals for both categories of TSD facilities are described below.

- **Interim Status TSD Facilities**

Groundwater regulations for interim status TSD facilities are set forth in Title 40 of the Code of Federal Regulations, Part 265, Subpart F. There are two stages of monitoring for interim facilities: detection and, if necessary, assessment monitoring.

For detection monitoring, the facility owner is required to monitor for three sets of parameters: (1) parameters characterizing the suitability of groundwater as a drinking water supply, (2) parameters establishing groundwater quality and (3) parameters used as indicators of groundwater contamination (indicator parameters). If comparisons of indicator parameters between upgradient and downgradient wells show a significant increase (or pH decrease), then the owner must resample and verify the results. If the results are still significant, the owner is required to move into the assessment monitoring phase.

During the assessment monitoring phase, the owner is required, at a minimum, to determine: "(i) The rate and extent of migration of the hazardous waste or hazardous waste constituents in the ground water; and (ii) The concentrations of the hazardous waste or hazardous waste constituents in the ground water." Based on these determinations, the owner must decide if hazardous waste or hazardous waste constituents from the facility have entered the groundwater.

The process of identifying groundwater quality information goals can begin by reviewing following statements given in Section 264 Subpart F of the RCRA regulations:

> "the owner ... must implement a ground water monitoring program capable of determining the facility's impact on the quality of ground water in the uppermost aquifer underlying the facility."
>
> The number, locations and depths of downgradient monitoring wells must ensure that "they immediately detect any statistically significant amounts of hazardous waste or hazardous waste constituents that migrate from the waste management area to the uppermost aquifer."
>
> The owner is required to determine "(i) The rate and extent of migration of the hazardous waste or hazardous waste constituents in the ground water; and (ii) The concentrations of the hazardous waste or hazardous waste constituents in the ground water."

- **Permitted TSD Facilities**

Groundwater regulations for permitted TSD facilities are set forth in Title 40 of the Code of Federal Regulations, Part 264, Subpart F. There are three stages of monitoring for permitted facilities: detection monitoring and, if necessary, compliance and corrective action monitoring.

For detection monitoring the owner is required to monitor for "indicator parameters, waste constituents, or reaction products that provide a reliable indication of the presence of hazardous constituents in ground water." The parameters or constituents to be monitored are specified in the facility permit. If the owner determines that there is statistically significant evidence of contamination for the specified chemical parameters or hazardous constituents, then he or she must "Immediately sample the ground water in all monitoring wells and determine whether constituents in the list of Appendix IX of Part 264 are present, and if so, in what concentration." If Appendix IX hazardous constituents are detected (a second analysis is permitted), then they will form the basis for compliance monitoring.

For compliance monitoring, the owner must continue to monitor for Appendix IX constituents and for each chemical parameter or hazardous constituent listed in the permit. In addition, the owner must determine whether regulated units are in compliance with the groundwater protection standard specified in the permit. If the groundwater protection standard is violated, then the owner must submit a "plan for a ground water monitoring program that will demonstrate the effectiveness of the corrective action."

A corrective action monitoring program must be as effective as the compliance monitoring program in determining compliance with the groundwater protection standard. In addition, it must be capable of "determining the success of a corrective action program."

The following statements given in Section 264 Subpart F of the RCRA regulations can be used to begin the process of identifying information goals:

> "The ground water monitoring system must ... yield ground water samples from the uppermost aquifer that (1) Represent the quality of background water that has not been affected by leakage from a regulated unit; (2) Represent the quality of ground water passing the point of compliance; (3) Allow for the detection of contamination when constituents have migrated from the waste management area to the uppermost aquifer."

> "The ground water monitoring program must ... ensure monitoring results that provide a reliable indication of ground-water quality below the waste management area."

"The sample size shall be as large as necessary to ensure with reasonable confidence that a contaminant release to ground water from a facility will be detected."

"Use of any of the following statistical methods must be protective of human health and the environment ... (1) A parametric analysis of variance (ANOVA) followed by multiple comparison procedures to identify statistically significant evidence of contamination."

"The owner or operator must monitor for indicator parameters (e.g., specific conductance, total organic carbon, or total organic halogen), waste constituents, or reaction products that provide a reliable indication of the presence of hazardous constituents in ground water."

"The owner or operator must determine whether there is statistically significant evidence of increased contamination at each monitoring well at the compliance point ..."

"... the owner or operator must establish and implement a ground water monitoring program to demonstrate the effectiveness of the corrective action program."

Comprehensive Environmental Response, Compensation and Liability Act (CERCLA)

General--The Comprehensive Environmental Response, Compensation and Liability Act (CERCLA or Superfund) established a program to identify, investigate, clean up and impose liability for abandoned hazardous waste sites (Cooke *et al.*, 1987b). CERCLA was passed on December 11, 1980 and was amended by the Superfund Amendments and Reauthorization Act (SARA) on October 17, 1986 (Hall *et al.*, 1990). One of the primary reasons for the enactment of Superfund was that RCRA does not authorize EPA to respond to toxic releases at abandoned hazardous waste sites (Glass, 1988). Except for the "imminent hazard" provisions, RCRA does not deal with abandoned facilities (Hall *et al.*, 1990).

The 1980 Superfund statute was a hastily drawn-up, compromise bill characterized by "numerous ambiguities, omissions, and poorly drafted provisions" (Cooke *et al.*, 1987b). A major controversy over the statute become known as the "How Clean is Clean?" question because Congress did not adequately address cleanup standards (Brown, 1990). The 1986 SARA reflects Congressional intent to resolve ambiguities in the 1980 statute. As well as modifying the existing legislation, SARA also added several provisions including a section on cleanup standards (Brown, 1990). The section on cleanup standards, however, is primarily narrative rather than quantitative and leaves EPA considerable discretion to fill in the details.

Groundwater Quality Information Goals Expressed in the CERCLA Statute--The CERCLA statute (as revised by SARA) may be found in the U.S. Code, Title 42, Sections 9601-9675. Section 9605 requires EPA to incorporate the provisions of SARA that relate to remedial action, into the National Contingency Plan (NCP). Specifically, CERCLA states that the NCP "...shall establish procedures and standards for responding to releases of hazardous substances, pollutants, and contaminants..."

Groundwater quality information goals are referred to in the following statements given in the CERCLA statute:

> "In assessing alternative remedial actions, the President shall, at a minimum, take into account: ... the goals, objectives, and requirements of the Solid Waste Disposal Act."

> "Such remedial action shall require a level or standard of control which at least attains Maximum Contaminant Level Goals established under the Safe Drinking Water Act ... and water quality criteria established under section 304 or 303 of the Clean Water Act (where relevant and appropriate)."

> In determining whether or not water quality criteria under the Clean Water Act is relevant and appropriate ..., the President shall "consider the designated or potential use of the surface or groundwater..."

As with RCRA, it is evident that EPA had limited guidance on what the information goals of groundwater quality monitoring should be.

Groundwater Quality Information Goals Expressed in the CERCLA Regulations--CERCLA implementation policy is codified by EPA in the National Contingency Plan (NCP) (U.S. EPA, 1988). The NCP specifies procedures, criteria and responsibilities for conducting response actions at Superfund sites (Cooke *et al.*, 1987b). The most recent revision of the NCP (March 8, 1990) can be found in Title 40 of the Code of Federal Regulations, Part 300.

In general, CERCLA does not contain the detailed type of regulations that are in RCRA. Instead, CERCLA uses broad terminology to outline the organizational structure and procedures for responding to releases. The procedures for responding to hazardous substance release will be briefly discussed in terms of the applications of groundwater quality monitoring.

Hazardous substance response is addressed in Subpart E of the NCP. There are two types of responses at Superfund sites: removal action or remedial action. Removal action refers to the "the cleanup or removal of released hazardous substances from the environment." Remedial action refers to a permanent remedy that is taken instead of, or in addition to, a removal action. Groundwater monitoring generally plays less of a role for removal action than it does for remedial action. If the decision has been made to implement removal action and there will be a planning period of at least six months, then a sampling and analysis plan is required for any environmental monitoring that is conducted. The plan "shall provide a process for obtaining data of sufficient quality and quantity to satisfy data needs."

The NCP outlines several steps in the remedial action process. The steps are as follows:

1. Remedial site evaluation.
2. Remedial investigation.
3. Feasibility study.
4. Remedy selection.
5. Remedial design/remedial action.

The remedial site evaluation involves data collection and evaluation of releases of hazardous substances, pollutants or contaminants. Although existing data is relied upon heavily at this stage, limited field sampling may also be necessary. If field sampling is conducted, a sampling and analysis plan is required.

The purpose of the next step, remedial investigation (RI), is to "determine the nature and extent of the problem presented by the release." Because data collection and site characterization are emphasized in the RI, groundwater monitoring often plays a large role.

The main objective of the feasibility study is to "develop and evaluate options for remedial action." The feasibility study should make use of data gathered in the remedial investigation. Information from the feasibility study is then used to select a remedial action that provides the best balance of trade-offs based on several criteria.

The remedial design results in a "detailed set of plans and specifications for implementation of the remedial action." A groundwater monitoring network design may be included in the plans. Once the remedial action is in operation, groundwater monitoring may be needed to assess the progress of remediation.

The following statements given in the NCP can be used to begin the process of identifying information goals:

> "The national goal of the remedy selection process is to select remedies that are protective of human health and the environment, that maintain protection over time, and that minimize untreated waste."

One of EPA's expectations that should be considered in developing appropriate remedial alternatives is to "return usable groundwaters to their beneficial uses wherever practical, within a time frame that is reasonable given the particular circumstances of the site."

"The purpose of the remedial investigation {RI} is to collect data necessary to adequately characterize the site for the purpose of developing and evaluating effective remedial alternatives."

"... a restoration activity will be considered administratively "complete" when: (i) Measures restore ground or surface water quality to a level that assures protection of human health and the environment;..."

MONITORING INFORMATION GOALS

Monitoring information goals are qualitative statements that describe specific information expectations of the monitoring program. They provide the underlying framework around which the data analysis protocol is organized.

Monitoring information goals are more specific than regulatory information goals. For example, a regulatory goal may be: **assess progress of remediation** and the corresponding monitoring goals may be: **determine the**

lateral extent of contamination plumes and determine the mass of contaminants removed from each remediation area.

Monitoring goals do not always have to correspond to a regulatory goal. In some cases, management may want information from the monitoring program that is not required by law. Statistical goals generally evolve from the monitoring goals as the means for obtaining the desired information are developed.

Although identification of monitoring information goals is site specific, there are a few underlying principles that are relevant in almost all situations. These are discussed here:

1. **Involve others in the formulation of monitoring information goals.**

   During the process of formulating information goals, the DAP writer should request input from others who are involved in the monitoring program. This may include regulators, managers, field samplers, chemists and environmental compliance staff. The chances for having a successful protocol are greatly increased if a consensus can be reached regarding the content of monitoring information goals.

2. **Don't include statistical terminology in the statement of monitoring information goals.**

There are two main reasons why statistics should not be included in the statement of monitoring information goals. First, monitoring information goals are intended to be clear, concise statements that can be easily understood and debated by anyone who is involved in the monitoring program. Many people are not comfortable with the language of statistics.

The second and most important reason to exclude statistical terminology is that statistics is a tool for achieving the monitoring information goal. It is not part of the goal itself. For example, a monitoring goal may be: **determine if concentrations of TCE at Well A are higher than background levels.** Statistics can be used to quantify the probability that a given concentration difference for a given sample size is due to chance. Some people, however, may find statistics helpful in further defining, precisely, what information will be produced.

3. **Be prepared to revise.**

Formulation of information goals is a dynamic process. Back and forth communication between individuals involved with the monitoring program will most likely result in changes to the originally drafted goals. Economics can also be a factor. Once the goals are identified, it may become evident that there is not enough money available to achieve all of them. Further revision of monitoring information goals may then be necessary.

STATISTICAL INFORMATION GOALS

Statistical information goals are complete, detailed statements that describe statistical intent. An example is: **detect monotonic, gradual trends in TCE concentrations equal to the standard deviation of the detrended data at the 95 percent significance level over a five year period of quarterly sampling.**

Statistical information goals for hypothesis tests should include three of the following parameters: (1) sample size, (2) the significance level, (3) the probability that a given difference will be detected (power) and (4) the magnitude of the differences. Assuming that the population standard deviation can be estimated, the above parameters are related in such a way that any three of them will define the forth.

The ideal way to write a data analysis protocol is prior to sample collection. If this is done, sampling size can be tailored to meet statistical information goals. The power can be determined for a particular sampling frequency, and if it is not acceptable, the frequency can be adjusted.

If the samples have already been collected, however, the protocol writer has limited control over power.

It is sometimes difficult to formulate a statistical information goal that closely matches the monitoring goal when using standard methods of hypothesis testing. For example, it is conventional for the null hypothesis to reflect the status quo of "no difference." The actual hypothesis of interest, however, may be whether a particular variable falls inside or outside a range of practical importance. Issues related to hypothesis testing are addressed in Chapter V.

Although the components of a statistical information goal will vary depending on the statistical method that is chosen, there are a couple of principals that are generally applicable. First, the statistical information goal should be specific and complete. Secondly, it should reflect the monitoring goal as closely as possible given the restrictions of the statistical method.

CONCLUSION

The importance of goal identification was addressed at the start of this chapter. Three types of information goals were then discussed: regulatory, monitoring and statistical. Regulatory information goals are the most difficult of the three to identify because the pertinent regulations are so vague and complex. Monitoring information goals, however, are the most important because they form the underlying framework of the data analysis protocol. Statistical information goals should be very specific.

CHAPTER IV

HANDLING ATTRIBUTES OF WATER QUALITY DATA RECORDS

INTRODUCTION

Handling of data record attributes is a major issue that should be addressed in data analysis protocols. Data record attributes arise from three factors: (1) statistical characteristics, (2) technological limitations and (3) operational failures.

Three statistical characteristics are discussed in this chapter: serial correlation, seasonality and distribution. One technological limitation, censoring, is addressed. Finally, several data record attributes that result from operational failures are reviewed briefly. These include missing values, different sampling frequencies, multiple observations and outliers. Chapters 3 and 6 of Ward *et al.*, (1990) provide background information on statistics that complements this thesis chapter.

SERIAL CORRELATION

Introduction

Water quality monitoring is often established in a sequential fashion through time. Depending upon how the water quality process being sampled behaves, samples may be provide redundant information. This tendency is referred to as serial correlation.

Serial correlation is one of the most difficult issues that a protocol writer has to deal with. There is widespread uncertainty and controversy over how serial correlation should be handled when analyzing groundwater quality data. Although serial correlation has been studied extensively by theoretical statisticians, the practical application of this research to analyzing groundwater quality data is poorly developed.

Several aspects of serial correlation that are important from the standpoint of a protocol writer are discussed herein. The following topics are covered: (1) definition of serial correlation, (2) effects of serial correlation on statistical analysis, (3) characterization of serial correlation, (4) ways of dealing with serial correlation and (5) recommendations.

Definition

Serial correlation is generally thought of as redundancy of information between adjacent observations in a time series. It can also occur between every other observation, every third observation and so on.

Positive serial correlation is characterized by the tendency for high values to follow high values and low values to follow low values. Negative serial correlation occurs when high values consistently follow low values and low values follow high ones. Only positive serial correlation will be addressed in the following discussion because water quality data is much more likely to exhibit positive than negative serial correlation.

Effects of Serial Correlation

Estimation--Confidence intervals for estimates of long-term parameters will be wider if the data is serially correlated than if it is independent. Thus, a larger sample size is needed for correlated data than independent data to achieve the same level of precision (i.e., the same confidence interval width) (Ward *et al.*, 1990).

The phrase "long-term" refers to a characteristic that is estimated from data that covers only a small portion of the time period of interest (Loftis *et al.*, 1991a). For example, a person may want to estimate the annual mean concentration of selenium for the past forty years. If the process is assumed to be stationary, the long-term mean can be calculated by averaging samples taken at fixed intervals for a year.

The assumption of stationarity, however, is unrealistic for most water quality applications (Loftis *et al.*, 1991a). It implies that the mean, variance and level of serial correlation are the same for every year in the time series

(Gilbert, 1987). A way to avoid this assumption is to calculate short-term parameters.

The phrase "short-term" refers to a characteristic that is estimated from data that covers only the time period of interest (Loftis *et al.*, 1991a). For example, a person may be interested in the mean concentration of selenium for a particular year. The short-term mean annual concentration can be calculated by averaging samples taken at fixed intervals for that year. This estimate of the mean will have a narrower confidence interval when the data is serially correlated than when it is independent (Loftis *et al.*, 1991a).

In most references, a clear distinction is not made between long-term and short-term estimates of parameters. Sample size and variance equations that account for serial correlation are generally derived based on long-term estimates (for example, see Gilbert, 1987 p. 38). There are sample size equations that account for serial correlation when the information goal is to estimate short-term characteristics. These equations are rather complex and are not widely published (Loftis *et al.*, 1991a). The appropriateness of long-term *versus* short-term parameter estimates for describing water quality data is an area that needs further research.

Hypothesis Testing--Serial correlation can distort the results of any statistical test that assumes stochastic independence of observations. Most hypothesis tests, including the nonparametric versions, have this assumption.

The most commonly cited effect of positive serial correlation on hypothesis tests is inflation of the Type I error. In other words, for a true null hypothesis, tests will reject the hypothesis at a higher rate than the nominal significance level (usually set at 5 percent). The end result is that the number of "false positives" are higher than they should be.[1]

Characterizing Serial Correlation

Three facets of characterizing serial correlation are discussed here: (1) choosing a model, (2) calculating the magnitude of serial correlation and (3) testing for significance of serial correlation.[2] The simplest and most

---

[1] Inflation of the Type I error in the presence of positive serial correlation has been noted for several types of tests including: Mann-Kendall and Seasonal Kendall tests (Harcum, 1990); t, sign and Wilcoxon tests (El-Shaarawi and Damsleth, 1988); hypothesis tests based on rank correlation coefficients such as Spearman's rho and Kendall's tau (Keller-McNulty and McNulty, 1987); multiple comparison procedures (Pavur, 1988) and ANOVA (Scariano and Davenport, 1987).

[2] It is commonly recommended that systematic characteristics such as seasonality and trend be removed from the time series prior to examination of serial correlation (Berryman, 1988; Montgomery and Reckhow, 1984 and Phillips *et al.*, 1989). Careful thought should precede removal of systematic characteristics, however, so that valuable information is not discarded.

commonly used model for analyzing serial correlation in groundwater quality time series is the lag 1 Markov model or AR(1). The model accounts for redundancy in the random term of adjacent observations (Harris, 1988). Additional information on model fitting can be obtained from Box and Jenkins (1970).

If the Markov (AR) model is chosen, serial correlation magnitude can be quantified by estimating the serial correlation coefficients, $\alpha(k)$. $\alpha(k)$ ranges from 0 to 1 for positive correlation and from 0 to -1 for negative correlation (0 is no correlation). "k" refers to the lag.

Equations for estimating $\alpha(k)$ are given in Box and Jenkins (1970) and are applied to water quality data in Loftis *et al.* (1989), Montgomery *et al.*, (1987) and Sanders *et al.* (1983). Harris (1988) and Close (1989) investigated the accuracy of estimates of the lag 1 serial correlation coefficient. They both found the estimates to be too low. Harris (1988) tried using a procedure developed by Quenouille (1949) to correct for bias. She found the procedure to be unsatisfactory. Close (1989) tested a method presented in Wallis and O'Connell (1972) to correct for bias in small samples ($n<100$). Monte Carlo simulations showed that the bias correction greatly improves the estimate of $\alpha(1)$, particularly for small sample sizes (Close, 1989).

Two approaches for testing the significance of AR(k) serial correlation are described here. One way is to compare the estimated correlation coefficient to a confidence interval. If the estimate falls outside the interval, the correlation is significant (Harris *et al.*, 1987). A graphical analog to this procedure is the correlogram. A correlogram is a plot of $\alpha(k)$ *versus* k for all lags k. Confidence intervals are displayed as horizontal lines on the correlogram. (Phillips *et al.*, 1989).

Harris *et al.* (1987) examined the power of the "confidence interval approach" to detect significant correlation of the AR(1) type. They found the power to be very low for small sample sizes. For example, the power to detect serial correlation in an AR(1) process with $\alpha(1) = 0.3$ and $n = 20$ is only 0.26. Harris *et al.* (1987) concluded that "Roughly speaking, it is not likely that moderate amounts of serial correlation can be detected in quarterly ground-water data without at least 10 years of sampling." Consequently, serial correlation is often not accommodated in groundwater data analysis procedures.

Another method for determining the significance of lag 1 Markov serial correlation is the rank von Neumann test (Gilbert, 1987). Harris (1988) compared the empirical power of the rank von Neumann test to that of the confidence interval approach described above. She found the rank von Neumann test to be more powerful.

Ways of Dealing With Serial Correlation

Introduction--Three approaches for handling serial correlation are discussed in this section. The approaches are: (1) disregard it, (2) avoid it by sampling infrequently and (3) use adjusted tests.

Disregard it--One option for dealing with serial correlation is to acknowledge it exists but don't make any adjustments in statistical calculations to account for it. The implications of this approach were introduced in the section on "effects of serial correlation."

Recall that positive serial correlation will tend to inflate the Type I error, thereby resulting in "false positives." This is usually considered to be an undesirable situation. There may be certain instances, however, when using a test that "mistakes" serial correlation for trend may be appropriate.

It is important to realize that the difference between serial correlation and trend is really one of scale (Loftis *et al.*, 1991a). For example, a five year drift towards higher (or lower) values would probably be modeled as serial correlation if the period of record were 30 years, and as trend if the period of record were only five years. If short term trends (relative to the length of record) are of interest from a management point of view, it may be appropriate to "ignore" serial correlation and use an unadjusted test, accepting that some error has been introduced due to violation of the independence assumption.

Another consequence of ignoring serial correlation is that the confidence intervals on estimates of long-term parameters will be narrower than they should be. This is an undesirable situation because it may lead to overconfidence in the estimate. Conversely, the confidence intervals on estimates of short-term parameters will be wider than they should be.

Serial correlation also distorts sample size calculations. If the goal is to estimate long term characteristics, the consequences of ignoring serial correlation are underestimation of sample size. On the other hand, if the goal is to estimate short term characteristics, the presence of serial correlation will lead to overestimation of sample size. Although this is a safe scenario in that changes or differences will be easily detected, it can be expensive because unnecessary sampling and analysis is conducted.

Avoid it--A commonly used tactic for dealing with serial correlation is to avoid it by sampling so infrequently that serial correlation is insignificant.

Montgomery *et al.* (1987) analyzed 118 quarterly groundwater quality data records from sites in California, Colorado and Illinois. They found that 17 of the well records exhibited significant (at the 5 percent significance level) serial correlation. The general consensus seems to be that most serial correlation in groundwater quality data can be avoided by taking quarterly samples (Harcum, 1990; Sara and Gibbons, 1991 and Ward *et al.*, 1988).

The RCRA guidance document for statistical analysis of groundwater data (U.S. EPA, 1989) contains a chapter on how to choose sampling intervals using the Darcy equation. Sampling intervals based on the RCRA method range from daily to monthly. Although the intent of the method is to set sampling frequencies that will result in independent data, it has been shown in the literature that monthly groundwater samples can exhibit high degrees of serial correlation (Close, 1987). Given this difference, it is obvious that this is an area needing further research.

If data has already been collected, an approach similar to sampling infrequently is to collapse (i.e., combine) observations until they are no longer serially correlated. For example, monthly observations could be collapsed to quarterly values. Harcum (1990) recommends using the mean to collapse data that has normally distributed errors, and the median for data that has lognormally distributed errors.

One advantage of using collapsed data values is that it has a lower variance and therefore a higher precision than single values (Ward et al, 1988). For example, quarterly averages will be more precise than a single quarterly observation (Ward *et al.*, 1988). In general, this advantage does not outweigh the cost involved in collecting the extra data.

Use adjusted tests--Another alternative for dealing with serial correlation is to use adjusted tests. Three categories of adjusted tests are discussed here: (1) Lettenmaier (1976) developed a method for using the Mann-Whitney and Spearman's rho tests in the presence of lag 1 Markov serial correlation, (2) Hirsch and Slack (1984) modified the popular Seasonal Kendall test so that it could be used with serially correlated data and (3) extensions of the Seasonal Kendall test to multivariate serially correlated data are discussed in Lettenmaier (1988) and Loftis *et al.* (1991c).

Knowledge of the persistence structure of the time series is required in order to use any of these adjusted tests. This can present a problem because large data sets are needed to characterize serial correlation (see previous section on characterizing serial correlation). Existing data is sometimes used to estimate serial correlation.

- **Adjustments for the Mann-Whitney and Spearman's rho tests**

Montgomery and Reckhow (1984) describe Lettenmaier's method in a clear and concise manner. A summary of that description is given below:

Step 1. Calculate the test statistic using the Mann-Whitney test for linear trends or the Spearman's rho test for step trends.

Step 2. Calculate the modified critical level.

Step 3. Compare the modified critical level to the test statistic calculated in step 1. The null hypothesis is accepted or rejected under the same circumstances as for independent data.

The power can then be calculated using the concept of "equivalent independent sample size."[3]

Step 1. Calculate the equivalent independent sample size.

Step 2. Calculate the trend number.

Step 3. Calculate the power.

---

[3] Lettenmaier (1976) introduced the concept of "equivalent sample size" for water quality trend evaluation. [He based his research on work done by Bailey and Hammersley (1946)]. For a given serially correlated time series, one can calculate the equivalent independent sample size that will provide the same amount of information as the correlated time series. For example, if weekly samples are collected (52/year), the effective independent sample size may only be 30/year due to the presence of serial correlation.

A major drawback of the adjusted Mann-Whitney and Spearman's rho tests is that errors in estimation of the lag 1 serial correlation coefficient can have a large impact on test results. Because accurate characterization of serial correlation in water quality data is difficult (or sometimes impossible) to achieve, these adjusted tests should not be included in a data analysis protocol as routine procedures. They may be useful, however, as exploratory techniques in the initial stages of writing a DAP.

- **Modified Seasonal Kendall test**

Hirsch and Slack (1984) adjusted the Seasonal Kendall test so that it is robust against serial correlation. The modified test preserves the Type I error in the presence of serial correlation except in cases where the data records are short or have high levels of persistence (Harcum, 1990; Hirsch and Slack, 1984 and Loftis and Taylor, 1989). As mentioned previously, serial correlation causes most tests to have inflated significance levels (i.e., a high rate of false positives).

Researchers have found that the modified Seasonal Kendall test has low power in some situations. Harcum (1990) showed that if there is no serial correlation, the modified test has lower power than either the Mann-Kendall or Seasonal Kendall tests. He noted that this difference in power decreases with increasing record length. Hirsch and

Slack (1984) also found that the modified test is less powerful than the original Seasonal Kendall test when there is no serial correlation. Loftis and Taylor (1989) conducted extensive simulation studies on seven different trend detection tests. They concluded that the modified test is "much less powerful than the other tests except for very large trend magnitudes and/or long data records."

Recommendations made by several authors regarding the use of the modified Seasonal Kendall test are summarized below:

1. Hirsch and Slack (1984) recommend that the test be used for data that arise from a stationary ARMA(1,1) process, with AR parameter $\Phi \leq 0.6$ and a record length of at least 10 years of monthly data.
2. Harcum (1990) recommends that the modified test be used on the original (monthly noncollapsed) data when there are ten or more years of data and serial correlation. He found that the significance level is preserved for ten or more years of data for "all but the highest level of serial correlation." The highest level of serial correlation he used was $\alpha(1) = 0.8$, whereas the second highest was $\alpha(1) = 0.6$.
3. Loftis and Taylor (1989) do not recommend using the modified test for routine application unless the data records are very long ("say > 20 years of

quarterly data"). They point out that the modified Seasonal Kendall test may "ignore" trends of moderate magnitude and duration that may be important from a management point of view. For example, a 30 year record could contain several 5 year trends, but unless a long term trend is present, the null hypothesis would not be rejected by the modified Seasonal Kendall test. Instead, the modified test would attribute the 5 year trends to serial correlation.

- **Multivariate tests**

The use of multivariate testing methods to analyze water quality data is increasing. One reason for this increased popularity of multivariate methods is that more water quality parameters are being analyzed in response to stricter regulations. Many water quality parameters can be grouped together based on their chemical compositions. Examples of such groups are common ions, nutrients and trace metals.

A single multivariate test can be applied to a whole group of parameters rather than applying univariate tests to each individual parameter. Loftis *et al.* (1991b) recommend the use of multivariate rather than univariate tests because multivariate tests usually have superior power.

Lettenmaier (1988) and Loftis *et al.* (1991c) performed Monte Carlo simulations on three types of multivariate tests that are able to handle serial correlation. These tests are discussed below:

1. **Covariance Sum (CS) Test:** This is a multivariate extension of the univariate Seasonal Kendall test. If all trends are in the same direction, this is the most powerful of the three tests. The CS test has low power, however, when both positive and negative trends are present.
2. **Covariance Inversion (CI) Test:** This method is based on work by Dietz and Killeen (1981). Lettenmaier (1988) and Loftis *et al.* (1991c) showed that the CI test has very low power in most situations. Unlike the CS test, it is not negatively affected by trends of different signs.
3. **Covariance Eigenvalue (CE) Test:** The CE method was developed by Lettenmaier (1988) in an effort to improve the power of the CI test. Lettenmaier (1988) and Loftis *et al.* (1991c) demonstrated that the CE test generally has much better power than the CI test. The power of the CE test is not negatively affected by trends of different sign.

Recommendations

Despite the controversy and uncertainty that surrounds the issue of how to deal with serial correlation when analyzing groundwater quality data, it is possible to make a some general recommendations. These recommendations are as follows:

1. Unless there is enough evidence to justify using a more complex model, assume that serial correlation can be described by a lag 1 Markov model.

2. For sample sizes less than 100, adjust estimates of the correlation coefficient, $\alpha(1)$, for bias using the method demonstrated in Close (1989).

3. To determine the significance of AR(k) serial correlation, use a correlogram or compare the correlation coefficient to a confidence interval. Use the rank von Neumann test to determine the significance AR(1) serial correlation.

4. Be aware that serial correlation can occur even in quarterly data records resulting in a higher level of false positives than what is indicated by the nominal significance level.

5. State the time scale of interest when estimating parameters, conducting hypothesis tests and determining sample sizes. The importance of serial correlation is dependent on the time scale of interest.

6. Consider the possibility of sampling monthly at the beginning of a monitoring program in order to enable earlier characterization of serial correlation (Harris, 1988). Monthly values can be collapsed to quarterly values for the purposes of detecting differences or changes.

7. Don't use adjusted Mann-Whitney or Spearman's rho tests as routine procedures in a data analysis protocol.

8. Use the adjusted Seasonal Kendall test only for long data records with low to moderate serial correlation. Remember that the adjusted test will have lower power than most unadjusted tests when there is no serial correlation present. Also, be aware that the adjusted Seasonal Kendall test will tend to ignore short term trends that may be important from a management standpoint.

9. When serial correlation is present and it is appropriate to use multivariate tests, use the Covariance Sum method if trends are homogenous and the Covariance Eigenvalue method if trends are of different signs.

SEASONALITY

Introduction

Seasonality is an important data characteristic that needs to be addressed in groundwater quality DAPs. As with serial correlation, there are no clear-cut, generally accepted guidelines on how seasonality should be dealt with when analyzing water quality data. This uncertainty regarding seasonality is evident in the EPA document *Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities* (U.S. EPA, 1989). Rather than providing guidelines on how to deal with it, the EPA document refers the reader to a professional statistician if seasonality is suspected.[4]

Seasonality in groundwater quality variables can be caused by changes in several factors including infiltration from streams, application of fertilizers, percolation from storm events and irrigation practices (Montgomery *et al.*,

---

[4] The EPA document (U.S. EPA, 1989) does include tentative recommendations on how to correct for seasonality when using control charts. The recommendations are summarized, however, by referring the reader to a professional statistician.

1987). Seasonality is the most prevalent in data collected from shallow or highly permeable aquifers (Montgomery *et al.*, 1987).

Definition

Seasonality is the change in distribution of water quality variables that can be attributed to the time of year. A "season" can be any specified period of time but is generally one month (12 seasons per year) or three months (4 seasons per year). Seasonality may or may not occur as a consistent pattern.

Effects of Seasonality

Seasonality increases the variance of water quality data, thereby increasing the width of confidence intervals in estimation procedures and decreasing the power of hypothesis tests. The majority of research that addresses the effects of seasonality on statistical analysis of water quality data, does so in terms of trend analysis. Seasonality can be important, however, in any statistical procedure that assumes stationary central tendency and dispersion.

Detecting Seasonality

Several approaches to checking for the presence of seasonality in water quality data are listed here:

1. Visually inspect a plot of concentrations *versus* time (Montgomery *et al.*, 1987).
2. Look for annual cycles in a correlogram. Details on how to construct and interpret correlograms are given in Loftis *et al.*, (1989).
3. Determine if there are physical factors such as a shallow or permeable aquifer that could lead to seasonality (Ward and Loftis, 1989).
4. Construct a box and whiskers plot for each season (Montgomery *et al.*, 1987). If the boxes do not overlap, then seasonality is probably a major source of data variation (Ward and Loftis, 1989).
5. Group the data according to season and calculate the mean for each season. Then calculate the ratio of maximum to minimum mean. The higher the ratio, the greater the seasonality. This can also be done with the standard deviation. (Loftis *et al.*, 1989).
6. Group the data according to season and conduct a Kruskal-Wallis or ANOVA test (Montgomery *et al.*, 1987).

The above methods involve a great deal of subjectivity. For example, one person may examine a time series and say that seasonality is present whereas someone else may look at the same time series and conclude that there is no significant seasonality. Even hypothesis tests (option #6) provide questionable results regarding seasonality, primarily due to the effects of sample size.[5]

Ways of Dealing With Seasonality

Introduction--For both hypothesis testing and estimation, there are two commonly used approaches for dealing with seasonality in water quality data: (1) performing seasonal transformations prior to data analysis or (2) using procedures that account for seasonality.

Some authors recommend that the presence of seasonality be verified in each individual data record prior to application of either of the two approaches for dealing with seasonality. Other authors recommend that historical data records be examined and, if seasonality appears to be a significant factor, subsequent data records should be treated as being seasonal.

A selection of publications that address ways of dealing with seasonality are discussed in the following paragraphs. The article titled *Groundwater Quality: A Data Analysis Protocol* was chosen for review because it includes straightforward examples of the two main ways of dealing with seasonality. Also, the article demonstrates the philosophy of verifying the presence of seasonality in each

---

[5] See chapter V for information on the difficulties involved with using hypothesis tests.

data record to be analyzed. The remaining publications were chosen because they represent the current state-of-the-art.

Groundwater Quality: A Data Analysis Protocol (Ward *et al.*, 1988)--In their data analysis protocol, Ward *et al.* recommend that special methods for dealing with seasonality be applied only if the presence of seasonality has been verified in the data record to be analyzed. Recommendations incorporated into the protocol regarding ways of dealing with seasonality are summarized in Table IV-1.

Notice that for trend analysis and medians comparison of independent data, the authors recommend using deseasonalized data (i.e., seasonal transformations) in conjunction with methods that do not account for seasonality. For excursion analysis (i.e., testing for a shift in concentration over a short time frame), they recommend using the original data with methods that do account for seasonality. Another important item shown in Table IV-1 is that seasonality is not a factor for medians comparison of paired data.

Techniques of Trend Analysis for Monthly Water Quality Data (Hirsch *et al.*, 1982)--This article describes the development of the Seasonal Kendall test which is the most commonly used method for trend detection in seasonal water quality data. The Seasonal Kendall slope estimator, an

TABLE IV-1. Methods for dealing with seasonality (Ward *et al.*, 1988).

Type of Analysis	Testing Method	Seasonality	Data to Be Used
trend analysis	Kendall Tau	present	deseasonalized
		absent	original
medians comparison: independent data	Mann Whitney	present	deseasonalized
		absent	original
medians comparison: paired data	Wilcoxon signed rank	not applicable	original
excursion analysis	compute p by season	present	original
	compute overall p	absent	

estimator of trend magnitude for time series that display seasonality, is also described. Applications of the Seasonal Kendall test and the Seasonal Kendall slope estimator are documented in several papers.[6]

Hirsch *et al.* used Monte Carlo simulations to compare their newly developed Seasonal Kendall test to: (1) linear regression on the original data and (2) linear regression on deseasonalized data. They found that linear regression on the original data is the best method to use if the data is known to be normal and nonseasonal. If the data is shown via statistical tests to be normal and seasonal, the best method is linear regression on deseasonalized data. They concluded, however, that the Seasonal Kendall test is the best overall method for trend detection because water quality data is always nonnormally distributed and often seasonal. They also pointed out that tests for detecting nonnormality and seasonality are unsatisfactory particularly for small data sets. The Seasonal Kendall test has the added advantage over linear regression methods of being able to handle missing values and nondetects.

---

[6] Papers that document applications of the Seasonal Kendall test and the Seasonal Kendall slope estimator include Alexander and Smith (1988), Lanfear and Alexander (1990), Walker (1991), and Lettenmaier *et al.*, (1991).

An Evaluation of Trend Detection Techniques for Use in Water Quality Monitoring Programs (Loftis *et al.*, 1989)--Loftis *et al.* used Monte Carlo simulations to compare several methods of trend analysis, all of which handle seasonality in some way. Those methods are:

- Analysis of covariance (ANOCOV)
- Modified t-test
- Kendall-tau following removal of seasonal means
- Seasonal Kendall
- Seasonal Kendall with serial correlation correction[7]
- ANOCOV on ranks
- Modified "t" on ranks

Data records were created with different patterns and magnitudes of seasonality in both the mean and standard deviation. Records with no seasonality were also generated. Out of the seven candidate tests, the authors recommended ANOCOV on ranks and the Seasonal Kendall test because they appeared to have the highest power.

Loftis *et al.* noted that ANOCOV on ranks has the benefit of being insensitive to the pattern and magnitude of seasonal change in variance. Also, ANOCOV on ranks can be conducted by using any statistical program that is capable of doing multiple linear regression. Finally, the ANOCOV

---

[7] The Seasonal Kendall test with correction for serial correlation is discussed in the section on serial correlation in this chapter.

method can be improved, if desired, by adding covariates to achieve better power or to model trends more accurately.

The Seasonal Kendall test, however, has been used extensively to analyze water quality data. Also, in the presence of serial correlation, it performs better than ANOCOV on ranks. For these reasons, the authors concluded that they have a slight preference for the Seasonal Kendall test.

Multivariate Tests for Trend in Water Quality (Loftis *et al.*, 1991c)--Loftis *et al.* recently investigated the use of multivariate methods for analyzing seasonal, serially independent water quality data.[8] They used Monte Carlo simulations to compare the performance of the following tests applied to 10 and 20 year data records.

- A method based on work by Sen and Puri (1977) [SP].
- Multivariate analysis of variance [MANOVA].
- A modified version of the covariance eigenvalue test[9] that assumes independence between seasons (i.e., no serial correlation) [MCE].
- A modified version of the covariance inversion test[9] that assumes independence between seasons [MCI].

---

[8] Multivariate methods for serially correlated seasonal data are discussed in the section on serial correlation in this chapter.

[9] The covariance eigenvalue and covariance inversion tests are discussed in the section on serial correlation in this chapter.

The authors found that for normal errors, MANOVA had the highest power overall followed very closely by the SP test. The MCE and MCI tests trailed behind in terms of power.

For lognormal errors, the SP test performed as well or better than the MCE and MCI tests. MANOVA had the lowest power.

Multivariate Trend Testing of Lake Water Quality (Loftis *et al.*, 1991b)--Loftis *et al.* used Monte Carlo simulations to compare univariate and multivariate methods for detecting trends in seasonal water quality data. Only normal, serially independent errors were considered in the study.

The authors found that MANOVA and the MCI test were "generally more powerful than their univariate counterparts applied using the Bonferroni inequality."[10]
Recommendations given by Loftis *et al.* are as follows:

> "Based on these results we can make a very positive recommendation for general application of multivariate approaches... Of the two multivariate methods studied, we recommend the MCI test for routine applications because of its robust performance (as demonstrated

---

[10] The "univariate counterparts" are ANOVA and the Seasonal Kendall test respectively. The Bonferroni inequality is used to control the overall significance level by performing K univariate tests at a significance level of α/K for each test.

elsewhere for rank-based methods in general) and its simplicity."

Recommendations

Some general recommendations regarding the issue of how to deal with seasonality when analyzing groundwater quality data are presented here. The recommendations are based on the preceding discussion, personal judgement, and conversations with researchers and practitioners involved in groundwater data analysis.

1. Determine whether seasonality is significant by examining historical data records. Then, based on that examination, assume that all subsequent data records will either be seasonal or nonseasonal and write the data analysis protocol accordingly. The alternative to this approach is to evaluate each new data record individually for the presence of seasonality. This alternative has two major drawbacks: (1) the methods that are available to detect seasonality involve a great deal of subjectivity, particularly for small sample sizes and (2) evaluating each new data record individually is time consuming. In addition, comparability of statistical results is enhanced if the same statistical analysis technique is used

throughout the study. If data records are evaluated individually, this implies that at least two statistical methods will be employed, i.e., one that deals with seasonality and one that doesn't.

2. Avoid using seasonal transformations whenever possible. Instead, use procedures that deal with seasonality directly. Seasonal transformations change the original data and may remove important information. Also, it can be difficult to interpret estimations made from transformed data.

3. If the information goal is to compare distributions, design the sampling program so that data can be paired. Seasonality is not an issue for tests that use paired observations such as the Wilcoxon signed rank test. In addition, tests which used paired data are more powerful than those which use independent data.

4. Use the Seasonal Kendall test if seasonality is suspected and the information goal is to detect trend. The Seasonal Kendall test has been used extensively to analyze water quality data. Also, it has been shown to have equivalent or higher

power to alternative methods. The Seasonal Kendall test can easily handle missing, censored or tied data values.

5. Keep informed of new developments in the use multivariate procedures for trend analysis of seasonal water quality data. Recent studies on lake water quality data have shown multivariate methods to be superior to multiple applications of univariate methods.

CENSORING

Introduction

Censoring occurs when chemists or data users replace numerical test results with qualitative statements. When censoring occurs, it is generally due to lack of confidence in the numerical result and fear that the data may be misused.

Terminology surrounding censoring is inconsistent, confusing and controversial. There is a question of whether or not censoring should even occur because it complicates statistical analysis and causes information loss. Several methods are available, however, to statistically analyze censored data.

A data analysis protocol writer needs to address several items regarding censoring including:

- Should the laboratory be allowed to censor analytical results?
- If detection limits or codes are used for reporting data, how are they defined?
- If data is censored by the laboratory, what statistical methods should be used to analyze it?
- How should low-level data be reported by the data user?

Definition and Causes of Censoring

Censoring is the replacement of numerical laboratory measurements with qualitative explanations such as ND, <T, less than LOD, or U. In water quality analysis, censoring generally occurs at very low concentration levels where measurement reliability is in question.

Censoring can occur at two stages. The chemist can censor laboratory results in reports to the data user, and the data user can censor results in reports to management or regulatory agencies.

Lambert *et al.*, (1991) listed the following factors that can lead to censoring by the chemist:

- The signal produced by the pollutant is too small for the instrumentation to discriminate from background noise.

- The instrumentation registers a low signal, but the chemist decides that "unpolluted" environmental samples could give a similar signal.
- A signal is registered, but certain criteria that identify the compound are not met.
- The measurement lies below a threshold set by a client or laboratory.

Censoring by the data user generally arises out of concern that individual data values will be misinterpreted. For example, a result of 0.10 ppb of TCE may cause alarm even if is well below the method detection limit. Negative results may cause people to doubt the overall integrity of the monitoring program.

If a single analyte is censored at more than one level, the data set is said to be "multiply censored." Multiple censoring can occur when: (1) various analytical methods are used for different ranges of contaminant concentration, (2) the amount of sample dilution varies or (3) detection limits decrease over time due to improved technology (Millard and Deverel, 1988).

Definition of Detection Limits

Detection limits are boundaries set up by chemists based on criteria given to them by the data user. The criteria may include acceptable levels of type I and type II error. Detection limits help to describe the uncertainty associated with detecting low-level contaminants. They are an aid to interpreting the significance of single data values, whereas statistical analysis is more appropriate for interpreting a group of values.

Detection limits are the most useful if they are determined separately for each complete analytical protocol and each individual contaminant. A complete analytical protocol is a documented procedure that includes all steps in the measurement process beginning with sample preparation and ending with data presentation.

There is a great deal of disagreement among scientists on how detection limits should be determined, what they mean and how they should be used. The confusion is evident in a statement made by a committee that studied ACS and ASTM approaches to detection limits: "...attempts by our task force on low-level data to make a rigorous conceptual and statistical comparison of the approaches have been unsuccessful. Even similar terms are defined in different, non-comparable ways, and additional terms and concepts are used which are unique to each approach."[11]

Beneath all the confusion, however, lie some fairly straightforward principles. These principles are briefly presented here. The definition of detection limits is the primary topic which is addressed. The use of detection

---

[11] The main authority on the subject of confusion over detection limits appears to be Currie (1968 and 1988).

limits is discussed in further detail in the section on reporting of low-level data.

Most of the currently accepted definitions of detection limits have a statistical basis and can be placed into one of three categories: limits based on type I error, limits based on type II error, and "other" (Table IV-2). Frequently used definitions are presented in Table IV-3. ASTM definitions will be used to discuss limits based on type I and type II errors.

The criterion of detection (COD), which is based on type I error, is used to answer the question: **is the substance present?** One way to describe the COD is in terms of a hypothesis test:

H_o: the concentration is equal to zero

H_a: the concentration is greater than zero

acceptable type I error = α

critical value = COD

If an analytical result is greater than the COD, the substance is assumed to be present. If the result is less than the COD, there is not enough evidence to claim that the substance is present.

The location of the COD can be determined from the pdf of H_o, the standard deviation of the analytical process and the choice of α. If ASTM guidelines are followed, the pdf of H_o is assumed to be normal and the standard deviation of the process is determined by one of three analytical

TABLE IV-2. Limits for describing statistical properties of low-level data.

Organization	Limits Based on Type I Error	Limits Based on Type II Error	Other Limits	Additional References
ACS (Keith, 1991)	**LOD** - limit of detection, or **MDL** - method detection limit	**RDL** - reliable detection limit	**LOQ** - limit of quantitation	(Keith, 1983) (Porter and Ward, 1991)
ASTM (ASTM, 1990)	**COD** - criterion of detection	**LOD** - limit of detection		(Kirchmer, 1988) (Wilson, 1973)
EPA (CFR, 1991)	**MDL** - method detection limit		**PQL** - practical quantitation limit	(Glaser *et al.*, 1981) (Keith, 1991)
Private Consultants for Love Canal Study (Lambert *et al.*, 1991)	**C_o** - decision limit	**D_o** - detection limit		(Currie, 1968) (Currie, 1988)

TABLE IV-3. Definitions of limits for describing statistical properties of low-level data.

Limit	Definition	Comments
LOD*	The lowest concentration level that can be determined statistically different from a blank at a specified level of confidence (Keith, 1991).	Often set at 3σ.
RDL	The concentration level at which a detection decision is extremely likely to be made correctly (Keith, 1991).	Often set at 6σ. RDL = 2 LOD* if $\alpha = \beta$.
LOQ	The level above which quantitative results may be obtained with a specified degree of confidence (Keith, 1991).	Often set at 10σ.
COD	The minimum quantity (analytical result) which must be observed before it can be stated that a substance has been discerned with an acceptable probability that the statement is true (ASTM, 1990).	The COD must always be accompanied by the stated probability (ASTM, 1990).
LOD**	A concentration of twice the criterion of detection when it has been decided that the risk of making a Type II error is to be equal to a Type I error (ASTM, 1990).	LOD** = 2 COD if $\alpha = \beta$.
MDL	The minimum concentration of a substance that can be measured and reported with 99 percent confidence that the analyte concentration is greater than zero and is determined from analysis of a sample in a given matrix containing the analyte (CFR, 1991).	Based on variability of analyte response rather than blank response (Kirchmer, 1988).
PQL	The lowest level that can be reliably achieved within specified limits of precision and accuracy during routine operating conditions (Keith, 1991).	

* as defined by ACS, ** as defined by ASTM

approaches presented in the guidelines. The relationship between the COD, α and the pdf of H_o are shown in Figure IV-1.

The limit of detection (LOD), which is based on type II error according to the ASTM definition, is used to answer the question: **what is the lowest concentration that can be reliably detected at a specific significance level?** The LOD can also be described in terms of a hypothesis test:

H_o: the concentration is equal to zero

H_a: the concentration is greater than zero

acceptable type I error = α

acceptable type II error = β, where $1 - \beta$ = LOD

critical value = COD

The location of the LOD can be determined from the pdf of H_a, the standard deviation of the analytical process, the choice of β and the location of the COD. ASTM assumes that the pdf of H_a is normally distributed and that the standard deviation is independent of concentration.

The relationship between the COD, LOD, α and β is shown in Figure IV-2 for the case of LOD = COD, and in Figure IV-3 for the case of $\alpha = \beta$. Notice that when LOD = COD, β is very high (0.5), and when $\alpha = \beta$, LOD = 2 COD.

The third category of detection limits (i.e., other), includes unique terms such as the LOQ and PQL that don't fit well into either of the first two categories. Both terms are defined in Table IV-3.

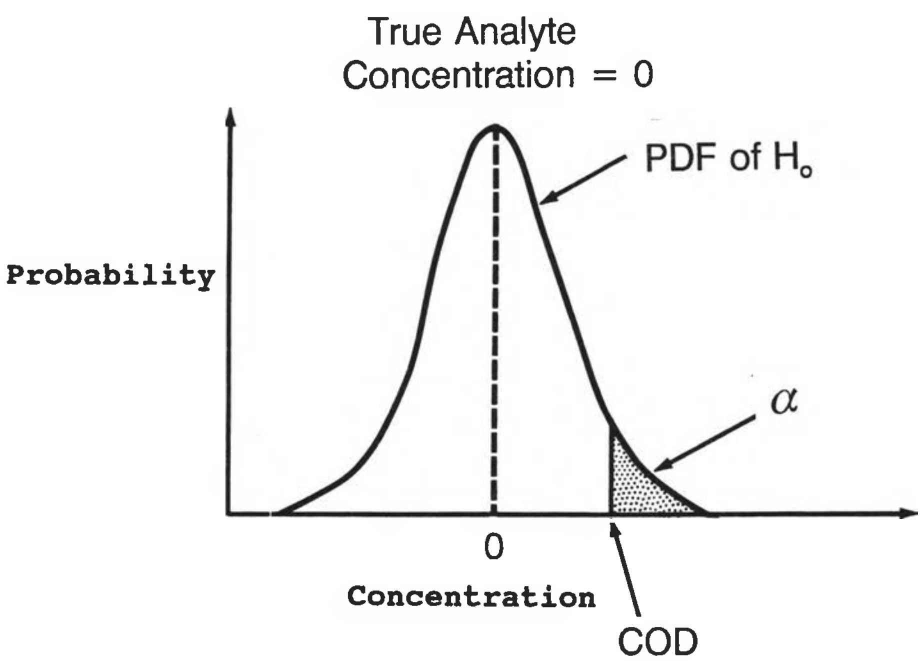


FIGURE IV-1. Graphic depiction of the relationship between the criterion of detection (COD), α and the PDF of H_o.

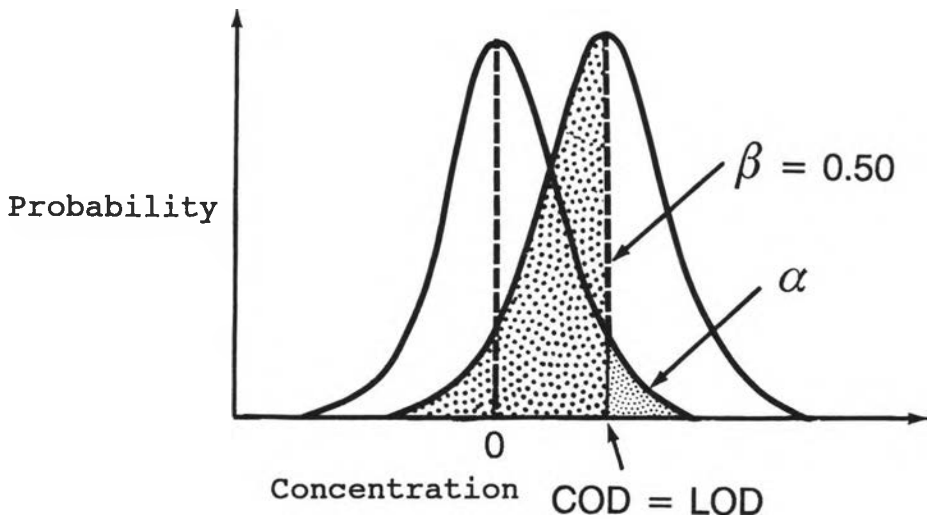


FIGURE IV-2. **Graphic depiction of the relationship between the criterion of detection (COD), α and β when COD = LOD.**

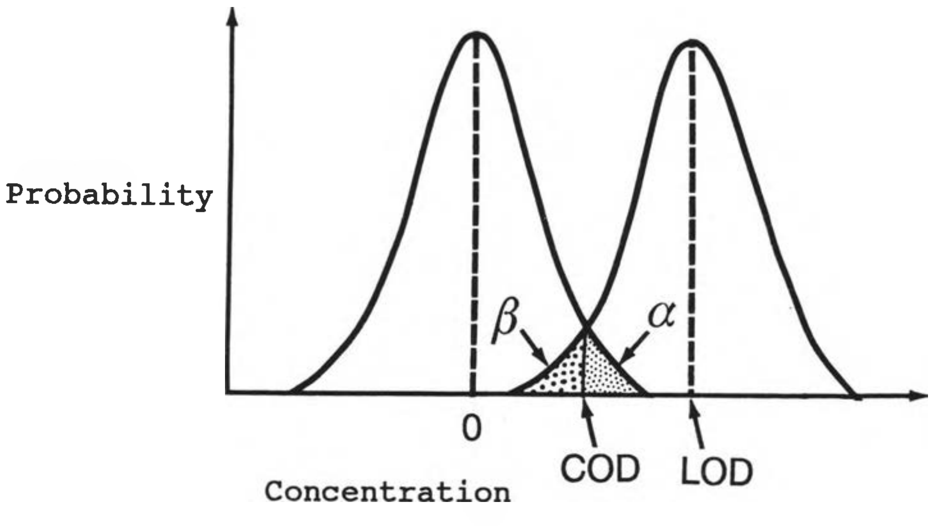


FIGURE IV-3. **Graphic depiction of the relationship between the criterion of detection (COD), limit of detection (LOD), α and β when α = β.**

The Effects of Censoring on Statistical Analysis

Information loss is the most commonly cited effect of censoring. Porter and Ward (1991) examined the loss of information caused by measurement noise and calibration, and compared it with information loss attributable to censoring. They found that, for the conditions they studied, uncensored samples provide more information about the central tendency of a parent distribution than censored data when censoring exceeds approximately 50 percent. Gilliom *et al.* (1984) examined the effects of censoring on trend detection capability. They concluded that: "For all classes of data evaluated, trends were most effectively detected in uncensored data as compared to censored data even when the data censored were highly unreliable. Thus, censoring data at any concentration level may eliminate valuable information."

The effects of censoring on statistical analysis of water quality data depend on: (1) the degree of censoring, (2) the statistical method being employed and (3) the quality of the data that has been censored. As the degree of censoring increases, the usefulness of methods for censored data (MCD) declines. For example, the power of tests to detect trend decreases as the percentage of censoring increases (Gilliom *et al.*, 1984). Also, when severe (near 50 percent or more) censoring occurs, hypothesis tests have little power to detect differences in central tendency (Helsel, 1990). Porter *et al.*, (1988) showed that confidence intervals for the estimates of the population mean get wider as the amount of censoring increases. Many of the MCDs simply cannot be used if censoring exceeds 50 percent (Helsel, 1990).

The effects of censoring on statistical analysis are also dependent on the statistical approach that is chosen. If simple substitution is used prior to estimation of summary statistics, the estimates may exhibit positive or negative bias depending on the substituted value. Substitution can also strongly influence the results of hypothesis tests. For example, a hypothesis test may declare a significant difference if NDs are replaced by zero, but no significant difference if NDs are replaced by the MDL. Deletion of censored values can cause hypothesis tests to have "little or no meaning" (Helsel, 1990).

The third factor that influences the effects of censored data on statistical analysis is the quality of the data (Gilliom *et al.*, 1984 and Taylor, 1988). If the analytical process used to generate the data was not in statistical control or if the bias is not predictable, then censoring may actually improve the results of statistical analysis. The obvious problem, however, is that factors which cause poor quality low-level data may also invalidate the higher concentration results.

Reporting Low-Level Data

Current recommendations that address how chemists should report their data are very much in favor of the no-censoring approach. Some of those recommendation are discussed here.

ASTM (1990) recommends that any instrument response below the COD (including negative values) be reported in conjunction with the "T" code. If no response is obtained, the "W" code should be reported along with the concentration that corresponds to the smallest increment that can be read on the analytical device.

In 1983, ACS recommended that, "Signals below 3σ should be reported as 'not detected' (ND) and the limit of detection should be given in parentheses" (Keith, 1983). In 1991, however, one of the major authors of the ACS guidelines effectively reversed this position by recommending that laboratories report all measurements (Keith, 1991). Keith also recommended that values below 3σ be flagged, and that the MDL or LOD be reported with all measurements.

The *RCRA Ground Water Monitoring Technical Enforcement Guidance Document* (U.S. EPA, 1985) states the following: "It is unacceptable to report only qualitative information values that were measured below a limit of detection. The technical reviewer must ensure that numerical values accompany the LT designation, so that the data are available for analysis."

The opinion that chemists should not censor data is also widely held among individuals. For example, see Lambert *et al.*, 1991; McNichols and Davis, 1988; Porter, 1986 and Porter *et al.*, 1988. The most widely held view seems to be that chemists should report measured concentration values along with a statement of uncertainty such as detection limits or confidence intervals.

Reporting of analytical results by data users is addressed by Keith (1991). He advocates the use of censoring if "the data user determines that the data may potentially be taken and used beyond the limits defining its measurement reliability."

Methods for Statistically Analyzing Censored Data

If the decision is made to request censored data from the laboratory, the data analysis protocol writer needs to become familiar with the various methods for censored data (MCDs). The purpose of this section is to provide a list of useful references on the subject.

Recommendations given by Helsel and Hirsch (1992) are summarized in Table IV-4. Methods for estimating summary statistics of censored data sets are discussed in the following references: El-Shaarawi, 1989; Gilbert, 1987; Harcum, 1990; Newman *et al.*, 1989 and Porter and Ward,

TABLE IV-4. Recommended methods for statistical analysis of censored data. Adapted from Helsel and Hirsch (1992).

APPLICATION		METHODS
Estimation of Summary Statistics	mean and standard deviation	• robust probability plot
	percentiles	• robust probability plot • MLE
Hypothesis Testing (single detection limit)	compare two groups	• rank sum test
	compare > two groups	• Kruskal-Wallis test
	> 50% censoring	• rank sum test • Kruskal-Wallis test • contingency tables
Hypothesis Testing (multiple detection limits)	compare two groups	• tobit regression
	compare > two groups	• tobit regression
Regression	< 20% censoring	• Kendall's robust line • tobit regression
	20% - 50% censoring	• tobit regression • logistic regression
	> 50% censoring	• logistic regression • contingency tables

1991. Methods for analyzing multiply censored data sets are provided in Hughes and Millard, 1988 and Millard and Deverel, 1988. Porter (1986) is an informative source for a variety of MCDs.

Recommendations

Recommendations regarding censoring are as follows:

1. Don't allow the laboratory to censor your data. Instead, have them report the measured concentration values along with statements of uncertainty such as detection limits or confidence intervals.

2. Use the uncensored results in statistical calculations.

3. If you feel that low-level data may be misinterpreted, consider using censoring in reports to management, regulators or the public.

4. If detection limits are used for any reason, make sure that both the determination and definition of the limits are well documented.

5. If MCDs are used, choose methods that are appropriate for the degree of censoring and the information goal, and robust to the assumption of normality. Also, be aware of the potential effects of censored data on the results of statistical analysis.

6. Keep informed of the latest developments on low-level data. It is a rapidly evolving field. A good place to start is to read Lambert *et al.*, 1991.

NONNORMALITY

Introduction

Water quality data are frequently right skewed and therefore violate the assumption of normality. Most parametric statistical methods which are used to analyze water quality data assume that the data are normally distributed. The power of such methods is sacrificed if the assumption of normality is violated.

To avoid the problem of loss of power due to nonnormal data, investigators are turning to nonparametric techniques. For some applications, however, nonparametric methods are not available. Also, regulations may require that parametric approaches be used. If parametric approaches are used, the distribution of the data should be investigated.

Several topics are addressed in this section on normality. The topics are: (1) distributional characteristics of water quality data, (2) effects of nonnormality on statistical analysis, (3) ways to deal with nonnormality and (4) recommendations for the protocol writer.

Distributional Characteristics of Water Quality Data

It is not uncommon for water quality data to be nonnormally distributed. They are often right (positive) skewed because they have a lower bound of zero with infrequent high values. The presence of nondetects accentuates right skewness.

Several investigations of the distributional properties of actual (*versus* simulated) water quality data records have been conducted. Some of these studies are summarized here.

Montgomery *et al*. (1987) examined 172 groundwater quality records for normality.[12] They found that 106 (62 percent) of the records were nonnormally distributed. The nonnormal records tended to be right skewed with the degree of skewness varying widely.

Gilliom and Helsel (1986) examined the statistical characteristics of trace constituent concentrations in

---

[12] Montgomery *et al*., (1987) used frequency histograms, normal probability plots, chi-squared goodness-of-fit test, and the skewness test to examine data records for normality. They used a significant level of 5% for the two hypothesis tests.

samples collected at U.S. Geological Survey river water quality monitoring stations. For 482 uncensored data sets, sample skewness ranged from -0.8 to 5.2 with a median of 1.8. Only 6 percent of the values were negative. A normal distribution is, by definition, symmetric and therefore has a skewness of zero. Gilliom and Helsel's results show that water quality data are frequently right skewed.

Loftis *et al.* (1989) checked water quality records from lakes for normality.[13] They found that only about 20 percent of the records displayed significant nonnormality. They repeated their analysis after log transforming the data and also after removing quarterly means. (Log transformations can sometimes make data more normal, and removing quarterly means can decrease the influence of seasonality thereby increasing the power of tests to detect nonnormality.) Loftis *et al.* concluded that neither action had much effect on the percentage of records that appeared to be nonnormal.

Effects of Nonnormality on Statistical Analysis

Two negative effects can occur when significance tests that assume normality are applied to nonnormal data. The

---

[13] Loftis *et al.*, (1989) used two hypothesis tests to check for the presence of nonnormality. The null and alternative hypotheses for each test were: (1) H_o: skewness = 0 and H_a: skewness ≠ 0, and (2) H_o: kurtosis = 3.0 and H_a: kurtosis ≠ 0. Both tests were applied at significance levels of 10% and 2% (i.e., 5% and 1% for each tail).

first effect is a distortion of Type I error (Snedecor and Cochran, 1989). For example, if the nominal significance level is set at 0.05 and the null hypothesis is true, H_o should be rejected 5 percent of the time. If the assumption of normality is not satisfied, however, H_o may be rejected say 3.6 percent or 8.5 percent of the time. Inflated significance levels will increase the incidence of false positives.

When tests are said to be robust against nonnormality, it generally means that nonnormality does not cause much distortion of Type I error (Conover and Iman, 1976). For instance, because the nominal significance level of the t-test is preserved for large sample sizes, the t-test is often cited as being robust against nonnormality (Helsel, 1987). This statement ignores the second undesirable effect that can occur when the assumption of normality is not satisfied, i.e, lack of power.

The power of a test that assumes normality can be lowered if the data is nonnormally distributed (Conover and Iman, 1976 and Helsel, 1987). This effect can result in false negatives. Low power is a potentially serious problem because important changes or differences in water quality may go unrecognized.

Ways to Deal With Nonnormality

Potentially nonnormal data can be dealt with the following ways:

- Use nonparametric statistical methods.
- Test for normality and if the data is normal, use parametric methods.
- Test for normality and if the data is nonnormal, use transformations. Proceed with parametric analysis if the transformed data is normal.
- Do a sensitivity analysis to examine the effects of nonnormality on the test you want to use. If the effects are acceptable, proceed with the test.

The first alternative, using nonparametric methods, has become the most accepted approach to statistically analyzing water quality data. One reason that nonparametric methods are so popular is because they do not assume normality. (Other advantages are discussed in Chapter V.) Hollander and Wolfe (1973) made the following statement regarding the efficiency[14] of nonparametric methods:

> "More often than not, the nonparametric procedures are only slightly less efficient than their normal theory competitors when the underlying populations are normal (the home court of normal theory methods), and they can be mildly and wildly more

---

[14] Efficiency is defined by Bradley (1968) as "a relative term comparing the power of one test with that of a second test which acts as a standard of comparison..."

efficient than these competitors when the underlying populations are not normal."

Bradley (1968) and Hirsch *et al.* (1991) have made similar conclusions regarding the efficiency of nonparametric methods.

The normality of data sets should be investigated prior to conducting parametric tests[15]. If the data is shown to be nonnormal, transformations can be calculated and the data should reexamined for normality (Helsel, 1987). If the data is still nonnormal or if transformations are not desirable, a sensitivity analysis can be conducted to quantify the effects of violating the normality assumption (Hirsch and Slack, 1984).

There are at least two problems with tests for normality. One problem with the tests is that they have very low power for small (<30) sample sizes (Helsel and Hirsch, 1988). This means that only extreme cases of nonnormality will be detected if the sample size is small.

The second problem is that even if a test for normality has adequate power, acceptance of the null hypothesis does not prove that the data is normal (Helsel and Hirsch, 1992 and Montgomery and Reckhow, 1984). It merely indicates that there is not enough evidence to indicate nonnormality.

---

[15] Methods to test for normality are presented in Chapter V. They are also discussed in Harris *et al.*, 1987, and Helsel and Hirsch, 1992.

Transformations are generally not desirable for multiple data sets (Hirsch *et al.*, 1991). Each data set should be examined separately to choose the best transformation. This can be quite time consuming. Also, if several transformations are chosen, it can be difficult to compare statistical analysis results between data sets. Another drawback of transformations is that the transformed data must be tested for normality (Helsel, 1987). It is not safe to assume that transformed data is automatically normal and therefore acceptable for analysis by parametric methods.

A sensitivity analysis to quantify the effects of nonnormality on a particular parametric statistical method may be appropriate in some situations (Hirsch and Slack, 1984). Because this is such a time consuming approach, it is probably only suitable for analyzing very important data sets by methods that do not have a nonparametric alternative.

Recommendations

The best approach for dealing with potential nonnormality in groundwater quality data records is to use nonparametric statistical methods. In general, parametric methods should not be used unless the data can be shown to be normally distributed.

DATA RECORD ATTRIBUTES THAT RESULT FROM DIFFICULTIES IN THE MONITORING PROGRAM

Introduction

Data record attributes that result from difficulties in the monitoring program include missing values, changing sampling frequencies, multiple observations and certain outliers. Every effort should be made to prevent these types of attributes from occurring. Even the best planned and operated monitoring programs, however, have occasional problems. Methods for handling data record attributes that result from operational difficulties are discussed below.

Missing Values

Missing values can be either random or systematic (Harcum, 1990). Random missing values may occur due to factors such as equipment failure, misplaced samples or test results, inclement weather, employee illness or war. Systematic missing values are often weather dependent. For example, wells may consistently dry up during the summer or be inaccessible during the winter. Changing sampling frequencies also result in systematic missing values (Lettenmaier *et al.*, 1991).

Some statistical analysis techniques which require regularly spaced samples in time or equal sample sizes, cannot be applied to data records that have missing values. Nonparametric methods can usually accommodate random missing data. Systematic missing values, however, can present special problems (Lettenmaier *et al.*, 1991).

Replacing missing values with numerical estimates is not recommended due to possible bias of statistical results. For most situations, the best alternative is to use nonparametric methods that can accommodate missing data. If there is a large percentage of missing values in a data record or if the missing values are systematic, the best alternative may be to collapse the data prior to statistical analysis. This alternative is discussed in further detail below.

Harcum (1990) investigated the effect of missing values on trend detection. He addressed the following question: "How many missing values are needed in a data set before it is necessary to collapse the data from monthly to quarterly values?" Harcum made the following conclusions:

- Collapse monthly data to quarterly values if more than 50 percent of the monthly data are missing when applying the Mann-Kendall tau or Seasonal Kendall tau tests.
- Collapse monthly data to quarterly values if more than 40 percent of the monthly data are missing when applying the Seasonal Kendall tau test with correction for serial correlation.

- When there are only five years of record and more than 50 percent missing values, there is not a good alternative.

Harcum's recommendations are probably the most useful to someone who is attempting to analyze existing data. A data analysis protocol writer should not have to deal with such large percentages of missing values if the monitoring program is well planned and operated.

Changing Sampling Frequencies

A variety of factors can cause changing sampling frequencies. Examples of these factors are: (1) increased funding, (2) changing regulatory requirements due to employee turnover in the regulators' office, (3) modified management priorities resulting from the discovery of new contaminants and (4) loss of funding.

Statistical methods which require equally spaced sampling intervals are not directly applicable to data that has been collected with changing sampling frequencies. The alternatives are to collapse or to exclude data so that all observations are equally spaced. Collapsing data will cause the periods with higher sampling frequencies to have lower variances. Excluding data causes information loss.

Multiple Observations

Multiple observations occur when more than one analytical result are recorded for the same time period. This generally happens when replicate samples are collected for QA/QC purposes. QA/QC data should be stored and analyzed with the rest of the data. If a single value is needed for a statistical method, the multiple values should be averaged. QA/QC data should not be discarded unless there is solid evidence that it is invalid.

Outliers

Outliers are values that are obviously higher or lower than the majority of data. They can have one of three causes: (1) a measurement or recording error, (2) an observation from a population not similar to that of most of the data or (3) a rare event from a single skewed population (Helsel and Hirsch, 1992). Outliers that result from either of the last two causes are true observations and should not be discarded.

Erroneous observations may be the result of several factors including sample contamination, failure of laboratory equipment, mistakes by the chemist or sloppy data entry. Statistical methods can give misleading information when erroneous observations are included in the data analysis. If there is evidence to show that an outlier is an erroneous observation, it should be discarded. Otherwise, it should be retained and used in statistical analysis applications along with the other data (Harcum, 1990).

Recommendations

The best way to deal with data record attributes resulting from operational difficulties is to prevent them from occurring. Writing a data analysis protocol prior to startup of the monitoring program is the first step towards prevention. An aggressive QA/QC program is also essential.

CONCLUSIONS

Several attributes of water quality data records were reviewed in this chapter. Serial correlation is probably the most difficult to understand and deal with. First, the distinction between serial correlation and trend is poorly understood. Also, the effects of serial correlation on statistical analysis depend on whether the investigator is interested in short-term or long-term parameters, which is another area that is poorly defined. Finally, there are no methods of dealing with serial correlation that are entirely satisfactory.

Tests that effectively accommodate seasonal data have been developed. In general, it is better to use these methods than to attempt to remove seasonality.

Censoring of water quality data can cause problems with statistical analysis methods. These difficulties can be avoided by requesting the laboratory to not censor the data.

Water quality data is often nonnormally distributed. Tests that assume normality may have low power when applied to nonnormal data. A simple solution to this problem is to use nonparametric methods.

Data record attributes that result from operational difficulties can be largely prevented by careful planning (i.e., writing a DAP). An effective QA/QC program is also necessary.

CHAPTER V

CHOOSING AND INTERPRETING STATISTICAL METHODS FOR ANALYZING WATER QUALITY DATA

INTRODUCTION

A data analysis protocol writer needs to know how to choose, apply and interpret statistical data analysis methods. Application of specific statistical procedures is delineated in other sources and is not covered here. The primary focus is on choice and interpretation of statistical methods for the analysis of water quality data.

This chapter begins with an overview on the general aspects of choosing statistical methods. Separate sections are then devoted to graphical techniques, point estimation, interval estimation and hypothesis testing. Each section includes brief summaries of some of the more commonly used procedures for analyzing water quality data. In addition, the sections on interval estimation and hypothesis testing contain detailed discussions on interpretation of statistical results.

CHOOSING STATISTICAL ANALYSIS METHODS

Overview

Three elements that should be considered when choosing methods to statistically analyze water quality data are: (1) monitoring information goals, (2) data record attributes and (3) characteristics of the proposed data analysis method. All three elements are discussed below. The third topic, method characteristics, is given the most emphasis because it is not dealt with elsewhere in this thesis.

It should be emphasized that this discussion addresses only one form of data analysis, i.e., statistics. Depending on the monitoring information goal, however, other forms of data analysis such as physical modeling or simple calculation procedures may be more appropriate. In general, statistical data analysis uses the laws of probability in conjunction with information regarding the random nature of water quality variables to provide an understanding of current water quality conditions.

Monitoring Information Goals

The first step in choosing a statistical data analysis method is to decide on a general statistical approach that matches the monitoring information goal. Several possible statistical approaches for analyzing water quality data are listed here:

- Compare two dependent groups.

- Compare two independent groups.
- Compare more than two dependent samples.
- Compare more than two independent samples.
- Determine the correlation between two continuous variables.
- Examine the relationship between two continuous variables.
- Observe and/or quantify behavior over time.
- Compare categorical data.
- Examine the relationship between continuous and categorical data.
- Estimate population distributional characteristics (e.g., mean, standard deviation).
- Verify frequency distribution assumptions (e.g., normality, equal variances).
- Estimate the probability that a single data value comes from a specific population.
- Estimate the probability that an interval contains the population value.
- Examine dependence structures in a data record (e.g., seasonality, serial correlation).

Once a statistical approach is selected, a specific method can be chosen based on data record attributes and the characteristics of candidate methods.

Data Record Attributes

The importance of considering data record attributes when choosing methods to statistically analyze water quality data is widely recognized. For example, section 25-8-204.5 of the Colorado Water Quality Control Act states, "In establishing water quality standards using statistical methodologies or in requiring the use of statistical methodologies for permit or enforcement purposes, **statistical methodologies used must be based on assumptions that are compatible with the water quality data**" (CDOH, 1992).

Consideration of data record attributes when choosing a statistical analysis method can present a "chicken-and-egg" problem for the DAP writer. A major theme of DAPs is that data analysis methods are chosen prior to data collection. How then, can attributes be identified and used to choose statistical methods if the data has not yet been collected?

As mentioned in Chapter IV, data record attributes in water quality data arise from three factors: (1) statistical characteristics, (2) technological limitations and (3) operational failures. Those that arise from statistical characteristics (i.e., serial correlation, seasonality and nonnormality), can often be predicted from prior monitoring results at the same site or even at a similar site.

Another approach for dealing with attributes that arise from statistical characteristics is to write a conditional data analysis protocol. Two or more options could be provided for statistically analyzing monitoring data. After the data is collected, an option could then be chosen based on observed data characteristics. There are two problems with this approach: (1) subjectivity and complexity of the protocol is increased and (2) groundwater quality data is often difficult to characterize if sample sizes are small.

The main data record attribute that results from technological limitations is censoring. As mentioned previously, this attribute can be eliminated by requesting laboratory personnel to not censor data. Not all water quality managers, however, subscribe to the no-censoring approach. If that is the case, the DAP writer must choose statistical methods that can handle the expected proportion of censored data.

The third class of data record attributes, those that result from operational difficulties, can be largely avoided by using effective QA/QC procedures. The DAP writer should be able to predict the presence of these attributes by reviewing the QA/QC program and by observing recent data records.

Characteristics of the Data Analysis Method

Point Estimation--The performance of estimators (i.e., point estimates of population parameters) is often gauged by their precision and bias (Berthouex and Hau, 1991). Precision refers to variability. A precise estimator has the ability to estimate a population parameter that is very close to the true population value from just one sample (Zar, 1984). This property is particularly important in groundwater quality monitoring because repeated samplings are prohibitively expensive.

The second criterion that is commonly used to evaluate estimator performance is the degree of bias. Bias refers to inaccuracy caused by persistent error. If a large number of samples are collected from one population, estimates of population parameters will eventually converge to the true value if the estimator is unbiased. A biased estimator will be consistently too high or too low regardless of sample size. (Zar, 1984)

Hypothesis Tests--The performance of hypothesis tests is frequently gauged by true significance level and power. (e.g., Conover, 1976; Harcum, 1990 and Loftis *et al*., 1989). The true significance level is determined with simulation studies. Tests are conducted on a large number of data sequences obtained from a population that meets the criteria of the null hypothesis. For example, if the purpose of the test is to detect trend, the true significance level would be determined by testing samples from a population that has no trend. The true significance level is the percentage of times that the null hypothesis is rejected (Loftis and Taylor, 1989). It is compared to the nominal significance level which is preassigned to the test and is commonly equal to 5 percent. The most desirable situation occurs when the true and nominal significance levels are equal (Montgomery and Loftis, 1987). If they are not equal, the true significance level will be unknown in actual (non-simulated) conditions. Incorrect conclusions could therefore be reached.

Some authors believe that power is the most important indicator of performance (e.g., Helsel and Hirsch, 1988). Power is the ability to detect departures from the null hypothesis. Parametric procedures can have low power if they are applied to nonnormal data (Helsel, 1987). This loss of power is critical from a water quality standpoint because it may result in nondetection of important differences or changes.

The relative power of two test methods can be determined by simulation studies. Each test is applied to a large number of samples obtained from a population with a known departure from the null hypothesis. If test A has a significantly higher proportion of detects than test B, then test A is said to be the most powerful (Harcum, 1990). It is not valid to compare tests that have different true significance levels (Harcum, 1990).

Nonparametric Methods--Nonparametric procedures are frequently chosen to statistically analyze water quality data. Several characteristics of nonparametric methods that make them practical for water quality data analysis are listed here:

- They are more powerful than their parametric counterparts for data sets that are nonnormally distributed.[16]
- Nonparametric procedures are resistant to outliers.
- They can be used to analyze data records that contain some censored values.
- Nonparametric methods are usually easier to understand and apply than their nonparametric counterparts.
- They can handle missing values.
- They can be applied to ordinal and sometimes nominal data.

References that discuss the advantages of nonparametric methods include: Daniel (1991), Gibbons (1985), Helsel (1987) and Hollander and Wolfe (1973).

---

[16] Chapter IV includes a more detailed discussion on this topic.

Validity of Simulation Studies

Many recommendations presented in the literature regarding choice of statistical methods are based on simulation studies. These recommendations should be viewed with discretion for a couple of reasons. The researcher has total control over the conditions of the study, including such factors as distribution shape, type and magnitude of trends, and sample size. If conditions in nature do not match those of the simulation study (they never will exactly), then recommendations based on the study may not apply. Consequently, it might be advantageous for DAP writers to review the conditions of the original simulation study.

Also, the criteria that researchers use to evaluate the effectiveness of statistical methods may have flaws that are not yet realized. DAP writers should, therefore, keep informed of the latest advances in statistical analysis of water quality data and be prepared to revise their protocols accordingly.

Additional Considerations

Although information goals, data attributes and method characteristics are the most frequently cited considerations for choosing statistical methods, other factors may also be important. One factor that can influence the choice of statistical analysis methods is the time scale of interest.

This can be particularly critical if the data is serially correlated (Loftis *et al.*, 1991a). Chapter IV includes a detailed discussion of this issue.

Another factor is the number of data sets to be analyzed. If only a few data sets are to be analyzed, it may be appropriate to use sophisticated statistical techniques such as multiple linear regression. However, if a large number of data sets are expected, simpler methods may be preferable (Loftis *et al.*, 1987).

The importance of results can also play a role in choice of statistical methods. For example, if a general awareness of water quality behavior over time is all that is needed, intervention analysis would be inappropriate.

Economics, politics and regulatory requirements may also influence the choice of methods to statistically analyze water quality data. Any factors that influence the choice of statistical data analysis methods should be stated in the DAP.

GRAPHICAL METHODS

Introduction

Graphical methods are used by water quality data analysts for a variety of applications including the following:

1. A tool for choosing statistical analysis techniques.

2. A means to interpret statistical results.

3. A "stand-alone" data analysis method.

4. A format for presenting information.

Application #1 should not be included as a routine procedure in a data analysis protocol because an underlying philosophy of DAPs is that statistical methods should be chosen prior to data collection. A DAP writer could, however, use graphical techniques to understand the general characteristics of water quality data that has already been collected at the site. This would provide the writer with valuable information regarding the type of statistical analysis methods that should be specified in the DAP.

Applications #2, #3 and #4 are appropriate for use in data analysis protocols. The importance of application #2, using graphical methods to interpret statistical results, is emphasized in many texts. For example, Ward *et al.* (1990) assert: "In general, it is recommended that a first step in data analysis would be to look at data in a graphical format in order to gain an understanding of water quality behavior which can be used in interpreting statistical results." Helsel and Hirsch (1992) make a similar statement: "Computing statistical measures without looking at a plot is an invitation to misunderstanding data..."

In some cases, data can be analyzed by graphical techniques alone (i.e., application #3). Graphical methods are particularly useful for analyzing limited data records.

For example, a time series plot can be more informative and less misleading than a formal time series analysis if data records are short.

Application #4, using graphical methods to present information, will not be specifically addressed here. Almost any graphical method that is useful for analyzing data and interpreting results, however, can also be used to present information. Presentation graphics are specifically addressed in Helsel and Hirsch (1992, Chapter 16), Schmid (1983) and Tufte (1983 and 1990).

A few of the most effective and practical graphical techniques are discussed herein. The following topics are addressed for each method: (1) potential applications, (2) types of information produced, (3) benefits and drawbacks and (4) modifications. In most cases, the actual mechanics of constructing the graph are not given, but appropriate references are included. Graphical methods for analyzing multivariate data are not covered. References that address graphs for multivariate data include Hem (1985) and Helsel and Hirsch (1992).

Examples of completed graphs are provided for most techniques.[17] The same data set, magnesium concentrations (hypothetical), was used for Figures V-1 through V-5. The

---

[17] All graphs for the section on graphical methods were generated on Statgraphics v.5.

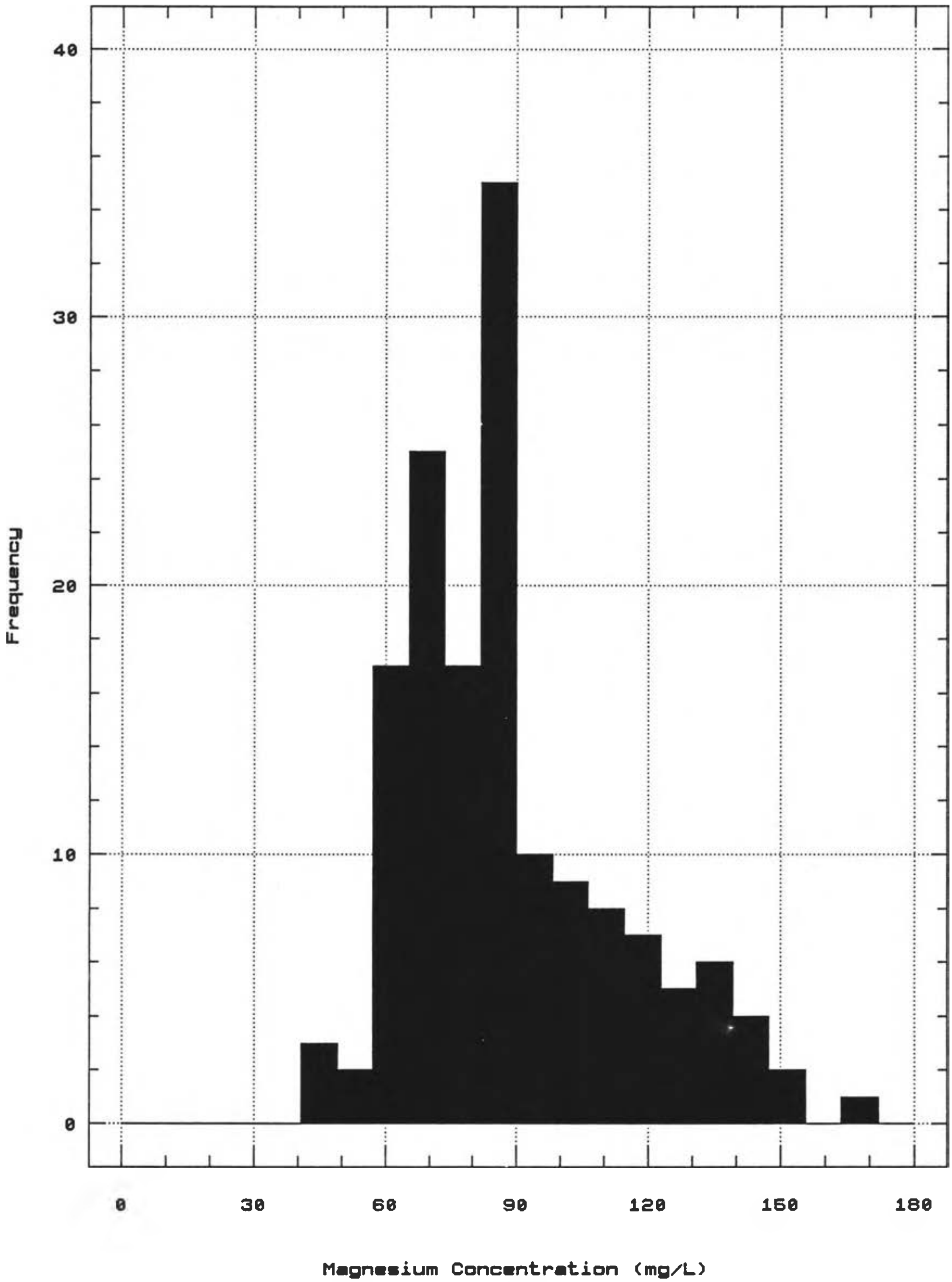


FIGURE V-1. Frequency histogram.

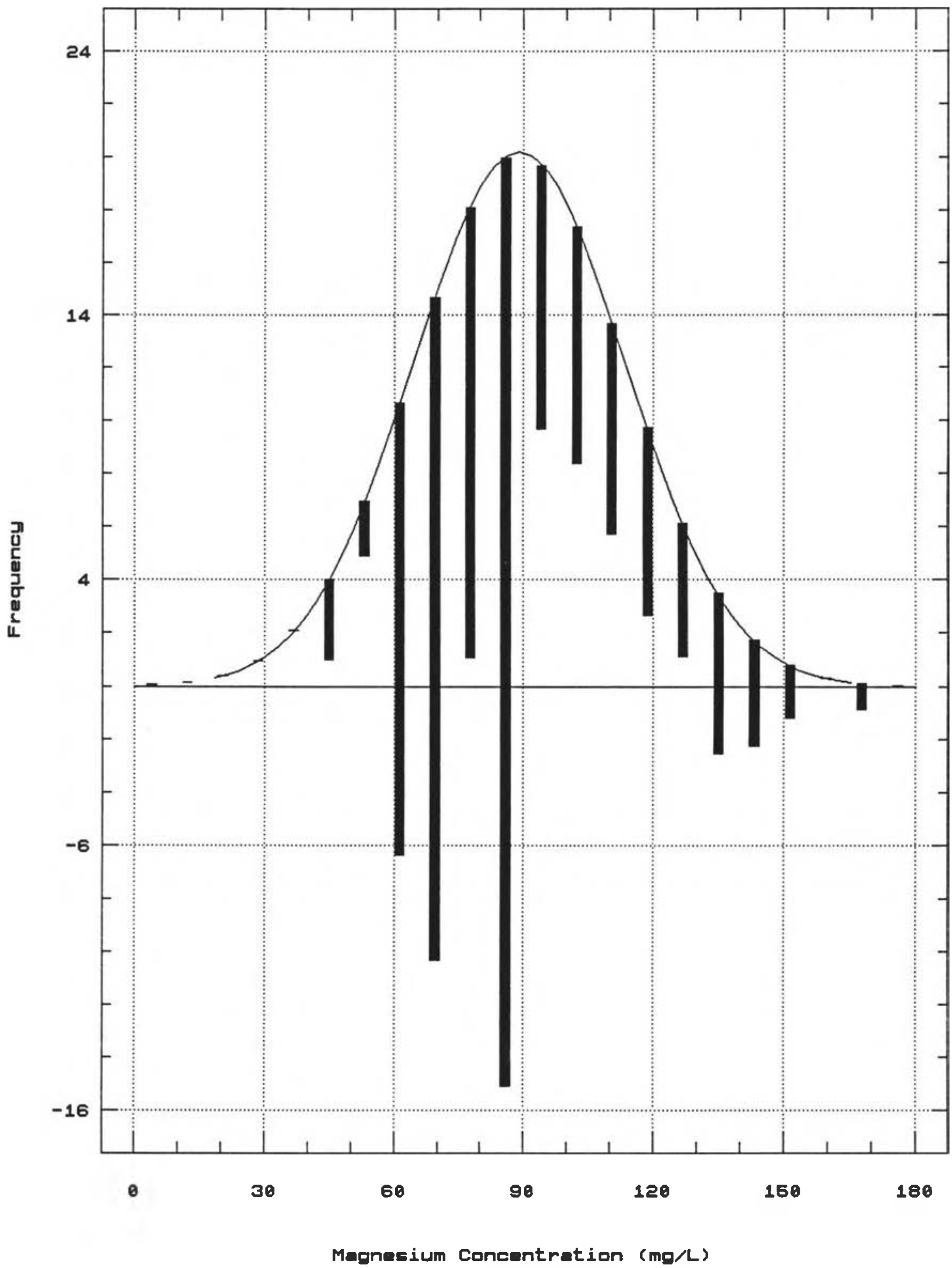


FIGURE V-2. Hanging histogram.

```
       unit = 1    1|2  represents 12

  3     4|888
  6     5|228
 37     6|0002233455555555677777777888889
 61     7|000000111244455555566789
(27)    8|000244444455555556888888888
 63     9|000000000002222555677
 42    10|000355555
 33    11|0000000255556
 20    12|005559
 14    13|0235899
  7    14|0025
  3    15|05

       HI|165
```

Magnesium Concentrations (mg/L)

FIGURE V-3. Stem and leaf diagram.

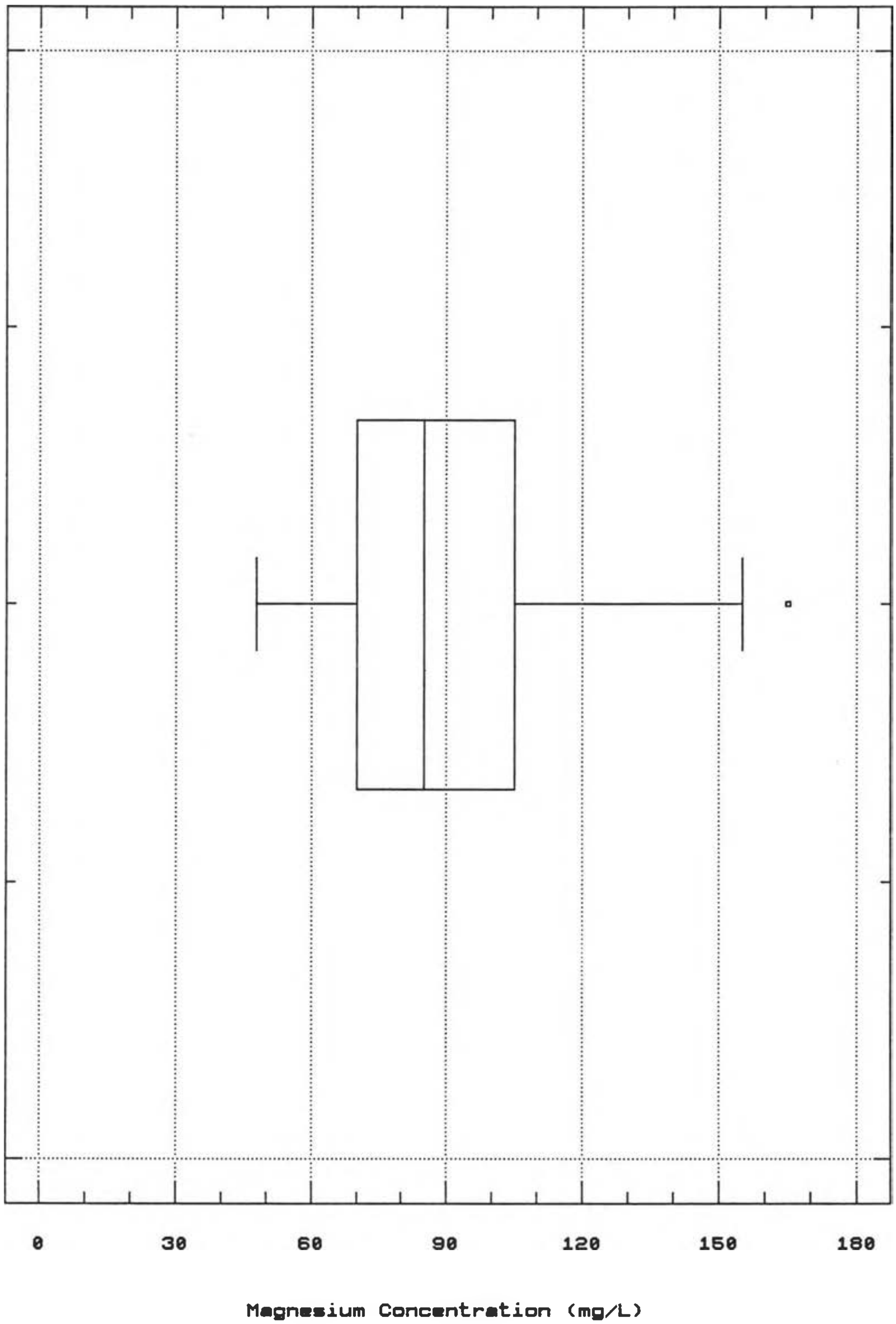


FIGURE V-4. Boxplot.

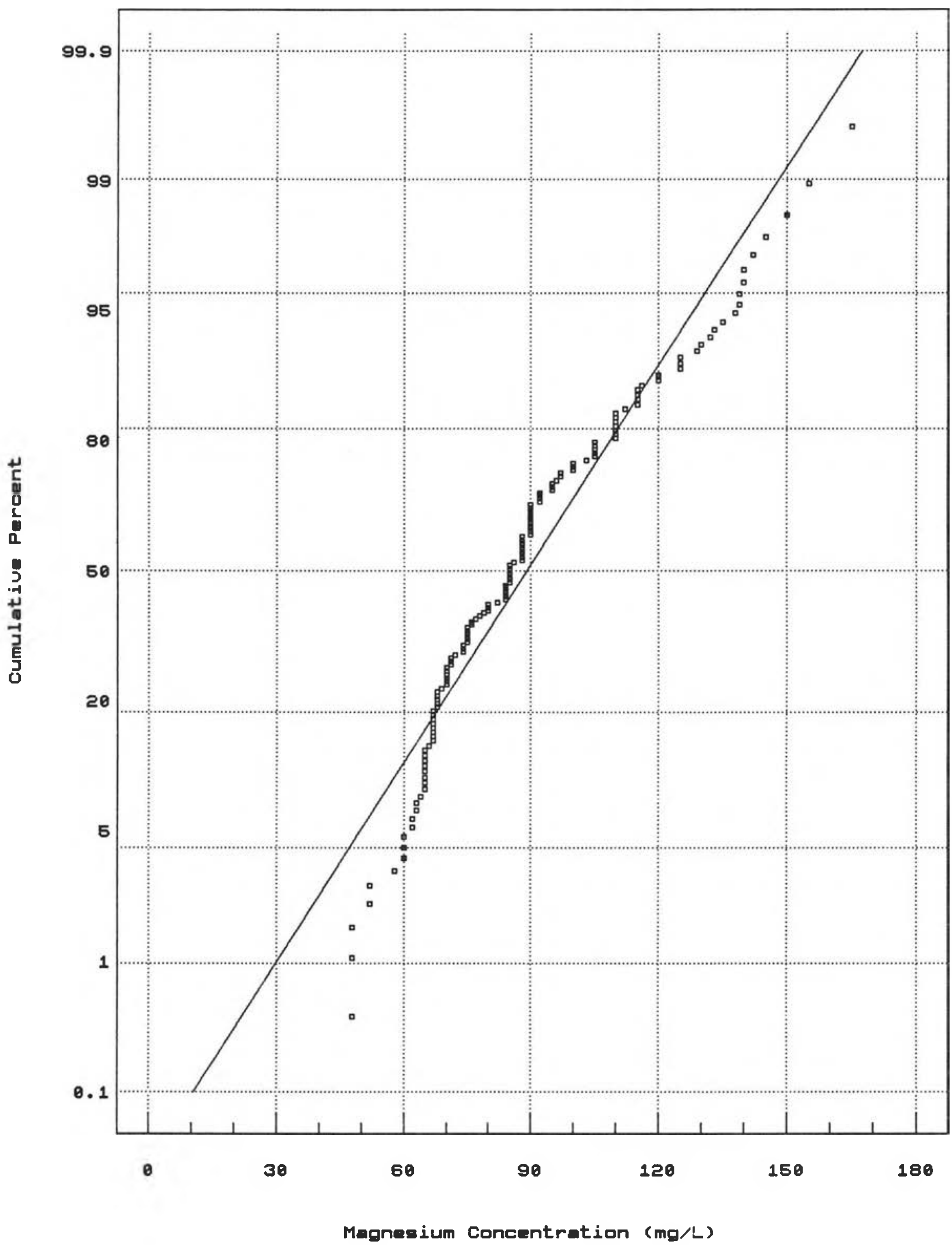


FIGURE V-5. Normal probability plot.

right-skewed nature of the data set is evident in all five figures.

Histograms

Histograms are a common graphical technique for displaying the frequency distribution of either discrete or continuous data.[18] Histograms consist of several vertical bars, usually of equal width and variable height. The area of each bar is proportional to the number or fraction of data points falling into one of several categories or intervals (Helsel and Hirsch, 1992). A typical histogram is shown in Figure V-1.

Histograms are useful for portraying the general shape and spread of sample distributions. They are not appropriate for detailed analyses.

Discrete data is better suited for analysis by histograms than is continuous data. Forcing continuous data into discrete groups can conceal important properties of the distribution. Furthermore, the visual impression of histograms constructed from continuous data is dependent upon the number of intervals that are chosen. Discrete data, however, often has natural groupings such as "the number of water-supply wells exceeding some critical yield grouped by geologic unit" (Helsel and Hirsch, 1992).

---

[18] Some authors use the term "histogram" for continuous data and "bar graphs" for discrete data (e.g., Weinberg and Goldberg, 1990).

Histograms are the most appropriate for observing a single set of data. They are too imprecise to be of much use for comparing multiple data sets. If they are used to compare two or more groups of data, however, they should be arranged one above the other rather than side-by-side or overlapping (Helsel and Hirsch, 1992).

A variation of the histogram is the rootogram which was developed by Tukey in 1972 (Wainer and Thissen, 1981). The areas of the vertical bars in a rootogram are proportional to the square root of the counts rather than the counts themselves because the roots are usually better behaved statistically (Wainer and Thissen, 1981). Both histograms and rootograms can be hung from the best-fitting normal distribution rather than plotted from the horizontal axis (Statgraphics, 1991 and Waiver and Thissen, 1981). If the data are normally distributed, the bottoms of the bars will be randomly scattered closely about the horizontal axis (Statgraphics, 1991). A hanging histogram is depicted in Figure V-2.

Stem and Leaf Diagrams

Stem and leaf diagrams are a method of displaying data that could be thought of as labeled histograms on their sides. They display more information than a histogram, however, because each individual data value is plotted. Stem and leaf diagrams are the most useful for analyzing single, small data sets. They can only be applied to continuous data. An example of a stem and leaf diagram is presented in Figure V-3.

Wainer and Thissen (1981) describe stem and leaf diagrams as "the most important device for the analysis of small batches of numbers to appear since the t-test." Stem and leaf diagrams allow the data analyst to do the following:

- visually examine the distributional properties of the data
- calculate the range of data values
- check for gaps and outliers
- compute order statistics

There are many different ways to construct stem and leaf diagrams. They are quite versatile and can easily be modified for a particular application. The most complete reference on constructing different types of stem and leaf diagrams is probably Tukey (1977). A simple explanation of how to produce a basic stem and leaf diagram is presented in Weinberg and Goldberg (1990).

Boxplots

Boxplots (often called box-and-whisker plots) provide a concise summary of several basic data characteristics. They are particularly useful for comparing changes in distributional characteristics of multiple data sets.

Because they are highly informative, yet simple to construct and interpret, boxplots have become a popular graphical method in many disciplines. They are widely used among water quality practitioners.[19]

Numerous variations of the boxplot have been developed.[20] The most commonly used version is probably the standard boxplot. A standard boxplot is shown in Figure V-4 and its construction is briefly described here. The description is based primarily on discussions by Helsel and Hirsch (1992) and Velleman and Hoaglin (1981).

The standard boxplot is divided into several regions. The box itself contains the center 50 percent of the data (i.e., the interquartile range). Assuming the box is constructed horizontally as shown in Figure V-4, the median is indicated as a vertical line within the box. The left end of the box is the 25th percentile and the right end is the 75th percentile. The inner fences are located at a distance of 1.5 times the width of the interquartile (IQ) range from either end of the box. The outer fences are located at a distance of 3.0 times the width of the IQ range

---

[19] For examples of actual applications of boxplots see Barton *et al.* (1987), Dubrovsky and Deverel (1989), Hirsch and Gilroy (1985), Howell and El-Shaarawi (1991), Kinnamon (1992), McLeod, *et al.* (1983), Montgomery *et al.* (1987), Newell *et al.* (1990) and Richards (1989).

[20] References that describe variations of the boxplot include: Becketti and Gould (1987), Benjamini (1988), Tukey (1977), Velleman and Hoaglin (1981) and Ward *et al.* (1990). Frigge *et al.* (1989) discuss the variety of definitions used for boxplots in statistical software packages.

from either end of the box. The fences are not actually marked on the boxplots, but are used to classify individual data points as described below.

Horizontal lines (i.e., whiskers) extend from either end of the box to the outermost value within the inner fence. Data points that fall between the inner and outer fence are called "outside values" and are individually plotted (Velleman and Hoaglin, 1981). Outside values occur fewer than one time in 100 for a normal distribution (Helsel and Hirsch, 1992). Data points that fall outside the outer fence are called "far outside" values and are plotted individually with a different symbol than what was used for the outside values (Velleman and Hoaglin, 1981). Far outside values occur less than once in 300,000 times if the distribution is normal (Helsel and Hirsch, 1992).

For a single data set, boxplots provide information regarding: (1) central tendency, (2) spread, (3) skewness and (4) the presence or absence of possible outliers. For multiple data sets, boxplots indicate how data distributions change over time and between locations.

Probability Plots

Probability plots are used to observe the goodness-of-fit between a set of data and a theoretical distribution such as the normal, lognormal or gamma distributions (Helsel and Hirsch, 1992). Although similar information can be obtained from histograms, deviations from a straight line are easier to see than deviations from a curved line (Helsel and Hirsch, 1992).

Probability plots can be produced by plotting quantiles of the sample data against quantiles of the standardized theoretical distribution. Alternatively, they can be plotted on specialized paper such as probability paper. Statgraphics follows the "probability paper" approach (see Figure V-5). Details of both types of probability plot construction are presented in Helsel and Hirsch (1992).

Probability plots are most commonly used to compare data to the normal distribution. Departures from normality show up as particular patterns on a normal probability plot. If the plot is arranged so that the horizontal axis represents the variable of interest as shown in Figure V-5, the following rule of thumb can be used: outliers at the left side of the plot will tend to fall above the linear pattern of the data, and outliers at the right side of the plot will tend to fall below the linear pattern. (The opposite is true if axes are arranged as shown in Helsel and Hirsch, 1992).

Based on this rule of thumb, the following interpretations can be made:

- Right skewed data will have a concave pattern as shown in Figure V-5.
- Left skewed data will have a convex pattern.

- Data with heavy tails[21] will have an "S" shaped pattern.
- Data with light tails will have a reversed "S" shaped pattern.

These interpretations are in general agreement with du Toit *et al*. (1986).

Normal probability plots are the graphic analog to the probability plot correlation coefficient test (Helsel and Hirsch, 1992). The two techniques are quite effective when used in conjunction with one another. The Shapiro-Wilk test is also related to probability plots (Helsel and Hirsch, 1992).

Time Series Plots

Time series plots are simply graphs of time *versus* a variable such as nitrate concentration. They allow the data analyst to visually examine the time series for seasonal variation, trends in all or part of the data, presence of extreme values, homogeneity of data, completeness of the data record and, to a certain extent, serial correlation (McBride and Loftis, 1991 and Phillips *et al*., 1989).

Time series plots are generally used in conjunction with quantitative statistical techniques. Quantitative techniques, however, can be misleading if applied to short

---

[21] "Heavy tails" refers to the occurrence of more data points in the tails than would be expected for a normal distribution.

data records. Time series plots are an excellent "stand alone" approach for analyzing short data records, particularly if data characteristics are poorly defined.

A technique known as smoothing can be used to eliminate some of the variability in the data, thereby making trends and cycles easier to detect. Smooths are intuitively appealing because they are calculated solely from the data. A model is not assumed. They are particularly useful for making sense out of large amounts of data.[22] Fox *et al.* (1990) used a type of smooth known as LOWESS to observe trends in freshwater inflow to San Francisco Bay from the Sacramento-San Joaquin Delta.

ESTIMATION

Point Estimation

A point estimate is a single number "best guess" of some characteristic of the population. A function that is used to obtain an estimate is called an estimator. Most point estimates can be transformed into interval estimates merely by adding confidence intervals.

Selected point estimators that are used for the analysis of water quality data are presented in Tables V-1, V-2 and V-3. Table V-1 includes estimators for distributional properties. Table V-2 summarizes estimators

---

[22] See Helsel and Hirsch (1992) and Tukey (1977) for additional information on smooths.

TABLE V-1. Estimators of distributional properties.

ESTIMATOR OR PROPERTY TO BE ESTIMATED	APPLICATION	COMMENTS	REFERENCE
mean	measure of location	Appropriate for computing units of mass.	Helsel and Hirsch, 1992
median	measure of location	Resistant to outliers.	" "
standard deviation	measure of spread	Unstable and inflated in the presence of outliers.	" "
interquartile range	measure of spread	Resistant to outliers.	" "
coefficient of skewness	measure of symmetry	Outliers can produce misleading results.	" "
quartile skew coefficient	measure of symmetry	Resistant to outliers. Uses only central 50 percent of data.	" "
quantiles		x_p, the *p*th quantile = that value below which lies 100*p*% of the population. Percentile = 100 x quantile.	Gilbert, 1987 Berthouex and Hau, 1991
proportions		p_{xc} = the proportion of the population that exceeds the value x_c.	Gilbert, 1987

TABLE V-2. Estimators of difference in location based on two samples.

ESTIMATOR OR PROPERTY TO BE ESTIMATED	APPLICATION	COMMENTS	REFERENCE
difference between mean values	independent samples	Used with the two-sample t-test unless data are transformed prior to testing. Then the Hodges-Lehmann estimator can be used.	Helsel and Hirsch, 1992
Hodges-Lehmann estimator	independent samples	The median of all possible pairwise differences between the two samples. Nonparametric. Used with the rank-sum test.	Hollander and Wolfe, 1973 case study - Chapter VII
mean difference	dependent* samples	Applicable where differences are symmetric and normally distributed. Used with the paired t-test.	Helsel and Hirsch, 1992
Hodges-Lehmann estimator	dependent* samples	The median of all allowable pairwise averages. Nonparametric. Used with the signed-rank test.	Hollander and Wolfe, 1973
median difference	dependent* samples	Used with the sign test. Nonparametric.	Helsel and Hirsch, 1992

* "Dependent" refers to samples that can be paired.

TABLE V-3. Estimators of linear dependence and monotonic trend.

ESTIMATOR OR PROPERTY TO BE ESTIMATED	APPLICATION	COMMENTS	REFERENCE
ordinary least squares (OLS) regression*	linear dependence	The following can be estimated: slope, intercept, y given x_i, residuals for observation *i* and mean square error. Assumes that residuals are normally distributed.	Helsel and Hirsch, 1992
Theil slope estimator*	linear dependence	More efficient than the OLS estimator when residuals are nonnormal. Nonparametric. Used with the Theil test for slope.	Helsel and Hirsch, 1992 Dietz, 1989 Hollander and Wolfe, 1973
Sen slope estimator	monotonic trend	Measures monotonic relationships. Nonparametric. Allows missing values. Used with the Mann-Kendall test.	Gilbert, 1987
Seasonal Kendall slope estimator	monotonic trend	Similar to the Sen slope estimator. Accounts for seasonality. Used with the Seasonal Kendall test.	Gilbert, 1987

*Can be used as a measure of trend magnitude if explanatory variable is time.

for the difference in location based on two samples. Estimators for correlation, linear dependence and trend are presented in Table V-3.

Hypothesis tests that are commonly used in conjunction with a specific estimator are identified in the tables. These comments are not meant to imply that the estimator has to be accompanied by a hypothesis test. Point estimates with confidence intervals (i.e., interval estimates) can be quite effective at conveying information. The relationship between hypothesis tests and estimation is discussed in subsequent sections.

Interval Estimation

Introduction--There are three main types of interval estimation procedures: (1) confidence intervals, (2) prediction intervals and (3) tolerance intervals. Definitions and applications of each approach are presented below along with references for further information.

Confidence Intervals--Confidence intervals are routinely used in the analysis of water quality data. A confidence interval is "a statement of the probability or likelihood that the interval contains the true population value" (Helsel and Hirsch, 1992). For example, a data analyst can be 95 percent confident that the true population parameter falls between the lower and upper limits of a 95 percent confidence interval.[23]

Confidence intervals are generally used in conjunction with point estimators. Equations for calculating confidence limits have been developed for most of the estimators listed in Tables V-1, V-2 and V-3. The width of the confidence interval indicates how much reliance should be placed on the point estimate. Similar information can be obtained from the yes/no results of a hypothesis test accompanied by the appropriate operating characteristic curve (Natrella, 1972). Confidence intervals, however, are much easier to construct and interpret than operating characteristic curves. Another important advantage is that confidence intervals, unlike operation characteristic curves, are presented in the same units as the original observations. (Natrella, 1972)

Confidence intervals are also used to detect outliers, for quality control charts, and for determining sample sizes necessary to achieve a stated level of precision (Helsel and Hirsch, 1992). Control charts are discussed in the case study (Chapter VII).

Tolerance Intervals and Prediction Intervals

Tolerance and prediction intervals are both used in groundwater data analysis to determine whether a new

---

[23] An excellent discussion on the interpretation of confidence intervals is presented in Jaeger (1990).

monitoring result is consistent with background monitoring levels. These two methods allow concentration values to be considered individually while still keeping the false positive rate at a reasonable level (Gibbons, 1991a).

Contrary to what is stated in the RCRA statistical guidance document (U.S. EPA, 1989), there are important differences between tolerance intervals and prediction intervals. These difference should be considered when deciding which of the two procedures to use.

Tolerance and prediction intervals can be either one-side or two-sided. For the purposes of this discussion, one-sided intervals will be assumed because they are appropriate for determining if a groundwater contaminant is higher than background levels. Except for pH, concentrations lower than background levels are generally not of concern. The following discussion is based primarily on a paper by Gibbons (1991a).

A tolerance interval states that a **given percentage of all future measurements** will fall in the interval with a specified level of confidence, if in fact, there is no difference from background levels. Therefore, a one-sided, 95 percent tolerance interval with 99 percent coverage means that a data analyst can be 95 percent confident that 99 percent of all future measurements in an uncontaminated well will fall in (or below) the interval.

There are two key points in the definition of tolerance intervals. First, tolerance intervals allow a specified number of false positives. In the example given above, the analyst could be 95 percent confident that 1 of the next 100 measurements will be a false positive. Secondly, tolerance intervals are independent of the future number of measurements. The "coverage" (99 percent in the above example) refers to a specified proportion of the entire population.

A prediction interval states that **all of a given number of future measurements** will fall in the interval with a specified level of confidence, if in fact, there is no difference from background levels. Therefore, a one-sided, 95 percent prediction interval constructed for the case of k = 20, means that the analyst can be 95 percent confident that all of the next 20 measurements from an uncontaminated well will fall in (or below) the interval.

There are also two important points in the definition of prediction intervals. The first is that no false positives are allowed. The price that is paid for this 100 percent coverage (i.e., no false positives) is an elevated false negative rate. In other words, important differences may not be detected. The second important point is that the width of the prediction interval depends on the number of future measurements for which it was constructed. The interval is narrower for fewer measurements. This feature makes prediction intervals appropriate for small monitoring programs.

Gibbons (1991a) developed a procedure using both tolerance and prediction limits that is particularly useful for situations where a large number of statistical comparisons are conducted on a regular basis. The procedure takes advantage of the strong points of both approaches and minimizes their weaknesses. Basically, he recommends that a tolerance interval be used initially, followed by the use of a prediction interval on results obtained from wells that failed the initial analysis.

Additional references on tolerance and prediction intervals that may be useful to the DAP writer include: Gibbons (1987a, 1987b, 1990, 1991b), Gibbons and Baker (1991), Helsel and Hirsch (1992), Loftis *et al*. (1987), Mee (1990) and Sara and Gibbons (1991).

HYPOTHESIS TESTING

Introduction

Hypothesis tests are used extensively in the analysis of water quality data. Like estimation, hypothesis tests provide a means for consistency. Each analyst that uses the same data and the same test will come up with identical answers. Unlike estimation, hypothesis tests provide a yes/no answer regarding the statistical significance of the evidence. The "p-value" is a measure of strength of that evidence.

Hypothesis testing has historically been the basis for statistical inference. Researchers from many disciplines, however, are beginning to question the value of hypothesis tests. A few authors recommend abandoning them altogether in favor of interval estimation procedures. The most common opinion, though, seems to be that hypothesis tests are important tools if they are applied correctly.

The following discussion is organized around the steps involved in conducting a hypothesis test. These steps are:

1. Choose the appropriate test.
2. Establish the null and alternative hypotheses.
3. Decide on an acceptable value for α.
4. Compute the test statistic from the data.
5. Compute the p-value.
6. Reject the null hypothesis if $p \le \alpha$.
7. Report the results.
8. Interpret the results.

Choose the Appropriate Test

The general principles involved in choosing a statistical analysis method were reviewed at the start of this chapter. Specific hypothesis tests that are used to analyze water quality data are presented in Tables V-4 thru V-8. Although the tables are not exhaustive, they do

TABLE V-4. Tests for goodness of fit.

TEST	APPLICATION	COMMENTS	REFERENCE
skewness test	normality	A simple test that works well for groundwater quality data analysis. Higher power than the chi-square test.	Harris *et al.*, 1987 Snedecor and Cochran, 1989
Lillefors test	normality	Similar to the Kolmogorov test. Test statistic can be found graphically.	Conover, 1980
Shapiro-Wilk test	normality	A powerful test that is related to the PPCC test and normal probability plot. More complex than the skewness test.	Conover, 1980 Harris *et al.*, 1987
probability plot correlation coefficient (PPCC) test	normality	Results can be visually illustrated on a probability plot, the graphic analog to the PPCC test.	Helsel and Hirsch, 1992
chi-square test	goodness-of-fit	Well known and versatile test. Less powerful at detecting nonnormality than most tests that are specifically designed for that purpose.	Harris *et al.*, 1987 Conover, 1980
Kolmogorov test	goodness-of-fit	Preferable to chi-square test when dealing with continuous data and/or small sample sizes.	Conover, 1980 Zar, 1984

TABLE V-5. Tests for location - two samples.

TEST	APPLICATION	COMMENTS	REFERENCE
t-test	independent samples	Properties have been widely studied. Less powerful than the rank-sum test when data are nonnormaly distributed.	Helsel and Hirsch, 1988 Zar, 1984
rank-sum test	independent samples	Also called the Wilcoxon rank sum, Mann-Whitney, or Wilcoxon-Mann-Whitney test. Nonparametric.	Conover, 1980 Hollander and Wolfe, 1973
paired t-test	dependent* samples	More powerful than the t-test when there is pairwise correlation. Assumes that differences are normally distributed.	Helsel and Hirsch, 1992 Zar, 1984
signed-rank test	dependent* samples	Assumes that the differences are symmetrically distributed. Nonparametric.	Conover, 1980 Hollander and Wolfe, 1973
sign test	dependent* samples	More generally applicable than the paired t-test or the signed-rank test. Easy to compute. Less powerful than the signed rank test. Nonparametric.	Conover, 1980 Gilbert, 1987

*"Dependent" refers to samples that can be paired.

TABLE V-6. Tests for location - more than two samples.

TEST	APPLICATION	COMMENTS	REFERENCE
one-factor ANOVA	independent samples	Widely used test. Assumes that all data groups have identical variances and are normally distributed.	Helsel and Hirsch, 1992 Zar, 1984
Kruskal-Wallis test	independent samples	Nonparametric equivalent of the one-factor ANOVA. Extension of the rank-sum test. Distributions of each data set are assumed to be identical in shape.	Hollander and Wolfe, 1973 Gilbert, 1987
two-factor ANOVA without replication	dependent* samples	The traditional parametric test for randomized complete block design.	Helsel and Hirsch, 1992 Zar, 1984
median aligned ranks ANOVA (MARA)	dependent* samples	An extension of the signed ranks test. Nonparametric.	Helsel and Hirsch, 1992
Friedman test	dependent* samples	The most common nonparametric test used for the randomized complete block design. An extension of the sign test.	Conover, 1980 Hollander and Wolfe, 1973
multiple comparisons procedures		Used only after an ANOVA-type test has been run that indicates a significant difference. Parametric and nonparametric versions are available.	Helsel and Hirsch, 1992

*"Dependent" refers to samples that can be paired.

TABLE V-7. Tests for linear dependence and monotonic trend.

TEST	APPLICATION	COMMENTS	REFERENCE
Ordinary least squares (OLS) regression*	linear dependence	The slope and intercept can be tested for departure from zero. Assumes that residuals are normally distributed.	Helsel and Hirsch, 1992
Theil test for slope*	linear dependence	Unlike the OLS, the Theil test is not strongly affected by outliers, and residuals can be nonnormaly distributed. Nonparametric.	Helsel and Hirsch, 1992 Hollander and Wolfe, 1973
Mann-Kendall test	monotonic trend	Missing values and non-detects are allowed. Nonparametric.	Gilbert, 1987 Helsel and Hirsch, 1992
Seasonal Kendall test	monotonic trend	An adaption of the Mann-Kendall test that accounts for seasonality. Widely used for water quality data analysis.	Gilbert, 1987 Hirsch *et al.*, 1982
multivariate trend tests	monotonic trend	Multivariate tests are more powerful than repeated applications of univariate tests. MANOVA best for independent data with normal errors and Sen-Puri test best for independent data with lognormal errors. CS test best for serially dependent data with homogenous trends. CE test best for serially dependent data with trends in both directions. See chapter IV.	Loftis *et al.*, 1991b and 1991c

*Can be used as a test for trend if explanatory variable is time.

TABLE V-8. Tests for correlation, equal variance and serial dependence.

TEST	APPLICATION	COMMENTS	REFERENCE
Pearson's r	correlation	Measures linear relationships. Not resistant to outliers. Assumes constant variance.	Helsel and Hirsch, 1992
Kendall's tau	correlation	Measures monotonic relationships. Nonparametric.	Conover, 1980 Helsel and Hirsch, 1992
Spearman's rho	correlation	Measures monotonic relationships. Nonparametric. Exact p-values should be used for $n < 20$. Spearman's ρ tends to be a larger number than Kendall's τ but the results of both tests are similar.	Conover, 1980 Helsel and Hirsch, 1992
F-test	equal variances	Severely affected by nonnormal distributions.	Zar, 1984
squared ranks test	equal variances	Nonparametric. Can be modified for use with more than two samples.	Conover, 1980
Rank von Neuman test	serial dependence	Tests for the presence of lag 1 serial correlation. Nonparametric.	Gilbert, 1987 Harris, 1988
confidence interval for r(k)	serial dependence	Lag 1 serial correlation is significant if r(1) falls outside the confidence interval.	Harris *et al.*, 1987

contain the majority of hypothesis tests that are routinely used for water quality data analysis.

Establish the Null and Alternative Hypotheses

The null hypothesis, H_o, is the hypothesis that is being tested. It is generally formulated to represent the status quo of no change or difference[24]. Some examples of typical null hypotheses are:

- The difference in means = 0.
- The correlation coefficient = 0.
- The slope coefficient = 0.
- The lag 1 serial correlation coefficient = 0.

The alternative hypothesis, H_a, represents a deviation from the status quo. H_a can be either one-sided or two-sided depending on the information goal. It is not valid to look at the data and then choose a one-sided or two-sided test. Typical alternative hypotheses are:

- The difference in means ≠ 0 (two-sided).
- The correlation coefficient ≠ 0 (two-sided).

---

[24] There has been some debate over the validity of formulating the null hypothesis to represent the status quo of no change or difference. Usually, the real hypothesis of interest in water quality is whether a particular variable falls inside or outside a range of practical importance (McBride *et al.*, 1992). A similar situation occurs in the area of bioequivalence testing. Several authors in that field have criticized traditional hypothesis testing methods and have suggested alternatives such as the equivalence test (Anderson and Hauck, 1983 and Patel and Gupta, 1984). McBride *et al.* (1992) discuss the use of equivalence tests to analyze water quality data.

- The slope coefficient > 0 (one-sided).
- The lag 1 serial correlation coefficient > 0 (one-sided).

If a one-sided test is chosen, the hypotheses should be written so as to be mutually exclusive. For example, if H_a is: the slope coefficient is > 0, then H_o should be: the slope coefficient is ≤ 0.

Decide on an Acceptable Value for α

The significance level, α, is the probability of making a Type I error. Type I error is the rejection of the null hypothesis when it is true. If the null hypothesis is true and α is set equal to 0.05, the null hypothesis will be rejected 5 times in 100 due to chance alone.

The data analyst is free to set α at whatever value they choose. The value of α, however, inversely affects the value of β, the Type II error. Type II error is the acceptance of the null hypothesis when it is false.

The choice of α should be based on the risk that management is willing to take of making Type I and Type II errors. A larger value of α will mean a larger number of false positives and a smaller number of false negatives.

Compute the Test Statistic from the Data

Calculation of the test statistic is dependent on the particular test that is chosen. There is usually only one way to calculate the test statistic for parametric procedures. Many nonparametric procedures, however, have three versions, all of which have different equations for determining the test statistic.

The three versions of nonparametric tests are: (1) exact test, (2) large sample approximation and (3) rank transformation test. The exact version produces exact rather than approximate results. The test statistic is compared to a table of quantiles which are calculated separately for every sample size or combination of sample sizes. Exact tests should be used for small sample sizes.

The large sample approximation can be used for large sample sizes (usually > 30). This test version assumes that the test statistic follows a particular distribution such as the normal distribution. The data analyst can compare the calculated test statistic with a table for the assumed distribution, thus eliminating the need for huge tables that include exact values for large sample sizes.

Most computer programs use large sample approximations for all sample sizes. This practice can introduce significant error. If sample sizes are small, the test statistic should be obtained from the program output and compared manually to a table of exact p-values.

The third version of nonparametric tests is the rank transformation test. Equations developed for parametric procedures are used to calculate the test statistic from ranks of the data. P-values are determined in the same manner as for the parametric test. The rank transformation version is useful in situations where a nonparametric analog is not available for a parametric test. This can occur if a nonparametric version has never been developed, such as for the multiple-factor ANOVA, or if the nonparametric version in not included in the statistical software that is being used.

Compute the P-value

The p-value is the smallest level of significance that would have allowed the null hypothesis to be rejected (Iman and Conover, 1983). P-values are obtained from tables and are based on the value of the test statistic.

Reject the Null Hypothesis if $p \leq \alpha$

Once the p-value is determined it should be compared to the significance level, α. Reject H_o if $p \leq \alpha$. Otherwise accept H_o.

Report the Results

The following information should be included when reporting the results of a hypothesis test:

- the name of the test that was used
- the chosen significance level
- whether H_o was accepted or rejected

- the p-value
- the sample size
- the data themselves or the location of the data

Some authors believe that an operating characteristic curve should also be included when reporting the results of a hypothesis test.

Interpret the Results

When interpreting the results of hypothesis tests, it is helpful to know that five parameters - sample size, significance level, power, the magnitude of the differences, and the population standard deviation - are interrelated in such a way that the establishment of any four of them determines the fifth (O'Brien and Shapiro, 1972). This information is also useful for designing a sampling program (Smith and McBride, 1990). The relationship between these five variables can be best understood by examining operating characteristic curves like the ones shown in Ward *et al.* (1990).

There are four possible results that can be obtained from hypothesis testing. These are shown in Table V-9 and discussed in the following paragraphs.

If the results of a hypothesis test indicate that there **is not** a significant difference, this can mean one of two things: (1) small differences exist and they are not

TABLE V-9. Decisions in hypothesis testing. Adapted from Iman and Conover (1983).

	Unknown True Situation	
	H_o is true (only small differences exist)*	**H_o is false (large differences exist)**
Fail to Reject H_o	**correct decision** probability of correct decision $= 1 - \alpha$	**Type II error** probability of Type II error $= \beta$
Reject H_o	**Type I error** probability of Type I error $= \alpha$ (significance level)	**correct decision** probability of correct decision $= 1 - \beta$ (power)

* H_o is never exactly true for continuous variables. Some difference always exists.

statistically significant or (2) large differences exist but they are not statistically significant. #1 represents a correct decision whereas #2 is incorrect. #2 usually occurs because the sample size is too small, the power of the test is too low, or the population standard deviation is too high.

On the other hand, if results indicate that there **is** a significant difference, this can mean: (1) large differences exist and they are statistically significant or (2) small differences exist but they are statistically significant. Again, #1 represents a correct decision whereas #2 is incorrect. #2 can occur if sample sizes are very large or the significance level has been set at a very low level.

Another key point pertains to statistical *versus* practical significance. Even if the hypothesis test is correct in declaring a large difference to be statistically significant, the difference is not necessarily significant or important from a water quality standpoint. This issue has been brought up by authors in a variety of disciplines including: pharmacy (Patel and Gupta, 1984), biology (Jones and Matloff, 1986 and Perry, 1986) and public health (Brown and Mikkelsen, 1990 and Ononoff and Boden, 1987).

The p-value is a critical component of hypothesis test results. Some authors recommend that only the p-value be reported so that the reader can make their own decision regarding significance. For two samples of the same size, a smaller p-value indicates a higher probability that the observed effect is "real" (i.e., that it can be attributed to something other than chance). The p-value also provides an indication of the confidence with which the null hypothesis is accepted or rejected. For example, if the p-value is only slightly larger or smaller than the significance level, the decision to accept or reject H_o is somewhat borderline.

CONCLUSIONS

The first step in a choosing statistical method should be to decide on a general statistical approach based on the monitoring information goal. Then, a particular statistical method should be chosen that is appropriate for the expected data record attributes. Information regarding the properties of estimators and hypothesis tests that are used to analyze water quality data is presented in Tables V-1 through V-8. Data should always be viewed graphically to clarify statistical results. Examples of practical graphical methods are shown in Figures V-1 through V-5.

CHAPTER VI

A FRAMEWORK FOR DEVELOPMENT OF GROUNDWATER QUALITY DATA ANALYSIS PROTOCOLS

INTRODUCTION

The framework presented in this chapter is essentially a "how to" manual for DAP writers. Although the framework was developed for hazardous waste facilities, it could be easily be modified for other situations such as municipal solid waste landfills. DAPs produced from the framework are to be written prior to sample collection. Existing data, however, can be used for characterization purposes or to confirm analytes.

The DAP design framework was developed with simplicity in mind. It is intended to be brief and easy to use. For this reason, detailed background material was relegated to Chapters III, IV and V.

DEFINE THE PROBLEM

The first step towards writing a data analysis protocol (DAP) is to define the problem. Review all existing information on site hydrology, geology and water quality. Become familiar with current remediation activities and regulatory requirements. In certain situations, a pilot study may be appropriate.

GAIN SUPPORT FOR THE PROTOCOL

The support of users, management and regulators is a critical requirement for a successful data analysis protocol. Gaining support for the protocol is an ongoing process that begins prior to protocol development. Possible approaches for obtaining support are as follows:

- Relate the advantages of a protocol. Mention that state and federal agencies, as well as private industry, are beginning to use water quality data analysis protocols.
- Involve others in the development process by arranging meetings and circulating draft copies of the protocol.
- Include the DAP in the company's audit process.
- Back up recommendations presented in the protocol with scientific evidence from the literature, and provide references for further information.

WRITE THE INTRODUCTION

The introduction sets the stage for the protocol. It should include purpose, scope, intended users and limitations.

INCLUDE PROCEDURES FOR DATA HANDLING

Data handling procedures address the computerization of data, data validation, and preparation of data for analysis. If the DAP writer can verify that data handling procedures are already well documented and routinely implemented, it may not be necessary to include them in the DAP. In most cases, however, data analysis procedures are poorly defined.

STATE INFORMATION GOALS

Identifying information goals may well be the most crucial step in writing a data analysis protocol. It is worth spending extra time on this task. The three types of information goals are: **regulatory, monitoring and statistical.** Details on identifying information goals are presented in Chapter III.

SPECIFY PROCEDURES FOR HANDLING DATA RECORD ATTRIBUTES

Viable options for dealing with data record attributes are presented below. Options that are preferable to the others for most applications are marked with an asterisk, *. Options that are unsatisfactory under any circumstances are not included. Further details are available in Chapter IV.

I. **Serial Correlation**

   A. Avoid it by sampling infrequently.

   B. Disregard it and be aware of the consequences (usually an increase in false positives).

C. Use methods that account for it.

II. **Seasonality**

A. Use methods that account for it.*

B. Pair data values.

C. Disregard it and be aware of the consequences (reduction of power).

III. **Censoring**

A. Request that data from the laboratory not be censored.*

B. If data is censored, use procedures that are robust to the assumption of normality, appropriate for the information goal and suitable for the percentage of censored values.

IV. **Nonnormality**

A. Use nonparametric procedures.*

B. Confirm normality prior to using parametric procedures.

C. Use transformations to obtain normality prior to using parametric procedures.

V. **Data Record Attributes Resulting from Difficulties in the Monitoring Program**

A. Avoid them by using an effective QA/QC program.*

B. **missing values**

1. Use methods that account for them.*

2. Collapse data.

C. **changing sampling frequencies**

   1. Collapse data.

   2. Exclude data.

D. **multiple observations**

   1. Average data.

   2. Discard only if evidence is available to show that the observation is erroneous.

E. **outliers**

   1. Discard only if evidence is available to show that the observation is erroneous.

CHOOSE DATA ANALYSIS METHODS

If statistical methods are chosen, base the choice on: (1) monitoring information goals, (2) data record attributes and (3) characteristics of the statistical method. This topic is addressed in detail in Chapter V. Examples of graphical techniques, tables of estimators and tables of hypothesis tests are provided.

DESCRIBE HOW STATISTICAL RESULTS SHOULD BE INTERPRETED

Always recommend that data be viewed graphically to aid in the interpretation process. Also, describe the basic information contribution of each statistical method. For example, if tolerance intervals are chosen, explain their information contribution as follows:

A tolerance interval states that a **given percentage of all future measurements** will fall in the interval with a specified level of confidence, if in fact, there is no difference from background levels.

If hypothesis tests are used, explain their limitations. It may help to state that: five parameters--sample size, significance level, power, the magnitude of the differences, and the population standard deviation--are interrelated in such a way that the establishment of any four of them determines the fifth. This relationship can be further defined by displaying an operating characteristic curve. The interpretation of statistical analysis results is further discussed in Chapter V.

SPECIFY REPORTING PROCEDURES

An effective way to specify reporting procedures in a DAP is to present actual samples of how the data is to be reported. This approach was used in the case study.

In many cases, it may be appropriate to recommend the use of graphical presentation methods. Many of the graphical analysis procedures presented in Chapter V can also be used for presentation.

If censoring terms, such as "less than MDL", are used to report data, the terms should be clearly defined in the report. When statistical analysis results are reported, the statistical method should always be identified. Details on how to report the results of hypothesis tests are presented in Chapter V.

DISCUSS HOW THE RESULTS OF STATISTICAL ANALYSIS SHOULD BE UTILIZED IN DECISION MAKING

Data analysis protocols should include explanations of how the results of statistical analyses will be used to achieve the information goals. Examples of these types of explanations are provided in the case study for information goals #2, #3 and #7.

It is important to emphasize in the information utilization section that statistical results should not be tied directly to decision making. Statistical results should be considered in conjunction with many other factors such as hydrogeology, flow patterns and contaminant transport behavior.

PREPARE A SUMMARY SHEET FOR EACH INFORMATION GOAL

A practical strategy for organizing data analysis protocols is to prepare a summary sheet for each information goal. This approach was very effective for the case study. Details that are specific to each information goal, such as sampling frequencies, wells to be sampled and compounds to be analyzed can be included in the summary sheets.

STATE PROCEDURES FOR REVIEWING PROTOCOL

Data analysis protocols should be reviewed periodically by the DAP writer (or a worthy successor) and revised if necessary. Procedures for reviewing the protocol could include:

- Period of time between reviews.
- Special situations that would warrant a review.
- Format for documenting revisions.
- Qualifications required of the reviewer.
- Factors to consider during the review such as current research, amended regulations and new information obtained from recent monitoring data.

INCLUDE ITEMS TO CLARIFY THE DATA ANALYSIS PROTOCOL

Numerous items can be added to data analysis protocols to enhance their clarity. A few suggestions are given here:

- Calculation procedures (see case study).
- Glossary of terms used in protocol.
- Flow chart showing organization of protocol.
- Procedures for utilizing existing data.

SUMMARY

The main subject headings for the DAP framework are listed in Figure VI-1. The figure allows the overall structure of the framework to be seen at a glance.

- DEFINE THE PROBLEM
- GAIN SUPPORT FOR THE PROTOCOL
- WRITE THE INTRODUCTION
- INCLUDE PROCEDURES FOR DATA HANDLING
- STATE INFORMATION GOALS
- SPECIFY PROCEDURES FOR HANDLING DATA RECORD ATTRIBUTES
- CHOOSE DATA ANALYSIS METHODS
- DESCRIBE HOW STATISTICAL RESULTS SHOULD BE INTERPRETED
- SPECIFY REPORTING PROCEDURES
- DISCUSS HOW THE RESULTS OF STATISTICAL ANALYSIS SHOULD BE UTILIZED IN DECISION MAKING
- PREPARE A SUMMARY SHEET FOR EACH INFORMATION GOAL
- STATE PROCEDURES FOR REVIEWING THE PROTOCOL
- INCLUDE ITEMS TO CLARIFY THE DATA ANALYSIS PROTOCOL

FIGURE VI-1. Main components of the framework for development of data analysis protocol for groundwater quality monitoring systems.

CHAPTER VII

APPLICATION OF THE DESIGN FRAMEWORK--A CASE STUDY

INTRODUCTION

The framework presented in Chapter VI has been used to develop a data analysis protocol for the IBM semiconductor manufacturing facility in Hopewell Junction[25], New York. The protocol will be included as an appendix to IBM's NYS (New York state) 373 permit that is currently being written. The development of a data analysis protocol represents part of IBM's effort to design and operate a state-of-the-art groundwater quality monitoring program.

Background information about the IBM East Fishkill site and their monitoring program is given immediately after this introduction, followed by a discussion section. The bulk of the case study is presented as appendices at the end of the thesis. The titles of the appendices are:

Appendix I: Initial submittal of groundwater monitoring proposal.

---

[25] Although the IBM facility is officially located in Hopewell Junction, it is known as the IBM East Fishkill facility.

Appendix II: Negotiations between NYSDEC and IBM regarding modifications to the initial groundwater monitoring proposal.

Appendix III: Final submittal of groundwater monitoring proposal.

Appendix IV: Data analysis protocol for groundwater monitoring conducted in response to NYS 373 permit--IBM East Fishkill Facility.

BACKGROUND

Site Description (Ward *et al.*, 1990)

The IBM facility is one of the largest semiconductor manufacturing plants in the world. The plant began operations in 1963. Currently, there are about 4 million square feet of buildings and 10,000 employees at the site. The site covers 750 acres in a semi-rural area in southeastern New York. It is surrounded on three sides by homes and a high school.

Approximately 2 million gallons of water are used per day at the IBM site. The water comes entirely from groundwater sources because there is no nearby municipal water supply. IBM treats all of its own water both before and after use.

Groundwater contamination problems at the site are due primarily to the presence of chlorinated hydrocarbons including Di-, Tri- and Tetra-chloroethenes. 1,1,2,2-

tetrachloroethene is the most commonly occurring compound. Chlorinated hydrocarbons are solvents that are heavily used in the manufacturing of semiconductors.

Most of the contaminants are assumed to have entered the groundwater between 1963 and 1978. Sources of the contaminants have been identified as leaks from buried solvent storage tanks and transporter pipes, spills associated with solvent handling, and seepage from a former construction debris landfill and two former fire brigade training areas.

IBM has virtually eliminated further groundwater contamination from the above sources. Solvents are stored in double walled tanks that are located above ground in concrete catch basins. Many of the pipes have been replaced with double-walled pipes on trestles. Solvent handling areas are underlain with concrete and fire training practices are environmentally safe. Also, 20,000 yards of material from the former construction debris landfill have been excavated and placed in a secure hazardous waste depository. The area is now filled with clean soils and capped with clay.

IBM currently operates a state-of-the-art groundwater remediation program. Contaminated water is pumped from the ground and treated with air stripping and/or granulated carbon. Approximately one million gallons of water are pumped and treated daily.

Monitoring Program

Groundwater at the IBM site has been monitored regularly since 1979. Major changes occurred in the monitoring system in 1982 and 1986. The initial monitoring program, the modifications in 1982 and 1986, and the monitoring system as it exists today, are briefly discussed in the following paragraphs.

The initial monitoring effort that began in 1979 produced data of low quality. The data record contained many missing values and non-detects. In addition, sampling frequency were inconsistent and data variability was high. The monitoring system design was not documented. Sampling and laboratory analysis was conducted by contractors.

IBM closely evaluated the monitoring system design and operation in 1982. The following modifications were made:

- IBM took over the field sampling and laboratory analysis.
- The monitoring system design was documented.
- Computerization of data was improved.
- Sampling points were grouped according to information goal (site perimeter, contaminated areas and general coverage).
- Emphasis was placed on analyzing compounds associated with IBM's prior activities at the site.
- Sampling frequencies were determined based on importance of location and past variability.

- A data verification protocol was implemented to catch outliers and confirm their validity.
- Reporting formats were specified.
- Quality control in the laboratory and field was given high priority.
- The monitoring system design document specified that data should be periodically evaluated to substantiate the original assumptions used in the design.

Additional refinements were made to the monitoring system in 1986 to increase the information content of the water quality data.

Although the 1982 modifications resulted in higher quality data with fewer missing values, more consistent sampling frequencies and less variability, the problem of nondetects still existed. IBM decided to follow the recommendations provided by Porter (1986) and ASTM D4210-83, and not censor their data. The laboratory was asked to report all the readings along with an estimate of the measurement error.

Data variability was further reduced at some of the key wells. Dedicated pumps were installed and a bladder system was used to seal off the portion of the aquifer that was being sampled.

Data analysis revealed that additional wells were needed at some locations. New wells were added to fill in knowledge gaps, to address new concerns and to monitor property that was added to the site.

The groundwater monitoring system today includes over 550 wells and piezometers. Approximately 100 of the wells are routinely sampled for water quality. More than 2000 samples are analyzed yearly for volatile organic compounds. Priority pollutant, inorganic, radiological and bacteriological analyses are conducted on some of the samples. Groundwater level measurements are taken in all monitoring wells and piezometers.

Regulatory Requirements

The site's groundwater monitoring and remediation activities have been conducted under an "Order-on-Consent" signed by IBM and the New York State Department of Environmental Conservation (NYSDEC) in 1981. IBM is currently in the process of obtaining a permit under Part 373, Title 6 of the New York Code of Rules and Regulations. The Part 373 permit, once granted, will regulate groundwater activities at the site.

Specifically, section 373-2.6.k.4 of 6NYCRR requires that "In conjunction with a corrective action program, the owner or operator must establish and implement a ground-water monitoring program to demonstrate the effectiveness of the corrective action program." The regulatory basis for

IBM's groundwater monitoring program is further explained in the data analysis protocol (i.e., Appendix IV).

DISCUSSION

The framework provided valuable guidance for writing a data analysis protocol for the IBM facility. The combination of flexibility in the basic framework and detailed background information was quite effective. It allowed the DAP to be site specific and scientifically defensible.

Use of the framework to develop the IBM groundwater quality data analysis protocol is briefly described in the following paragraphs. The discussion is structured around the main components of the DAP framework which are listed in Figure VI-1.

- **Define the Problem**

Extensive work has already been conducted at the IBM site to define the problem, so very little additional work was required for this first step.

- **Gain Support for the Protocol**

Support was gained for the protocol primarily by holding company meetings. The concept of a DAP was introduced and potential advantages were discussed. Many practical suggestions were made at the meetings.

Participants included field samplers, laboratory personnel, water quality staff and managers. NYSDEC was not involved in writing the DAP. Rough drafts of the protocol were circulated at various stages during the development process.

- **Write the Introduction**

Topics covered in the introduction include: purpose and advantages of the protocol, scope, intended users and format of the protocol. Limitations were not discussed in the introduction because they are addressed in other sections of the protocol.

- **Include Procedures for Data Handling**

Data handling procedures are well established at the IBM site. Additional guidance and documentation in this area is unnecessary. Data handling procedures were, therefore, not included in the data analysis protocol.

- **State Information Goals**

Regulatory goals were defined by reviewing legislation and by conversing with regulators. An outline of the applicable laws and regulations was included in the protocol to guide the reader through the legislative maze.

A considerable amount of effort was devoted to formulating monitoring information goals. The effort paid off. For one thing, the list of goals served as a synopsis of information that would be obtained from the monitoring program. The list provided an effective means to identify concerns that had been overlooked or improperly emphasized. In addition, monitoring information goals proved to be a convenient basis for organizing the protocol. Details of the protocol such as sampling frequency and data analysis method were specified on a summary sheet for each information goal. The goals were numbered so they could be easily referenced throughout the protocol.

Detailed statistical information goals were stated on the sample calculation sheets.

- **Specify Procedures for Handling Data Record Attributes**

Background information provided in the framework was very useful for writing this section of the protocol.

- **Choose Data Analysis Methods**

Statistical procedures were chosen for only four of the twelve monitoring information goals. (One of the four procedures was graphical.) This rather surprising lack of emphasis on statistics can be traced back to the concerns of the regulators. Regulators are primarily concerned that contaminated groundwater is contained, collected and treated. Direct comparison with fixed limits is often all that is needed to ensure that contaminated water is being contained. Pump readings and calculations of contaminant mass can provide information regarding collection and treatment.

Another reason for not using statistical methods is the concern that results may be misinterpreted by others. For example, upward trends in contaminant levels may be associated with unsuccessful remediation efforts, when in fact, they could actually be indicative of effective remediation.

In situations where statistical analysis methods were chosen, the recommendations provided by the framework were closely followed.

- **Describe How Statistical Results Should be Interpreted**

Interpretation of results was described for each statistical method. Difficulties related to interpretation of hypothesis tests were not an issue because estimation was chosen instead of hypothesis testing.

- **Specify Reporting Procedures**

Reporting procedures were specified by including samples of all reporting formats.

- **Discuss How the Results of Statistical Analysis Should be Utilized in Decision Making**

Information utilization was addressed primarily in the DAP summary sheets. It was also discussed in Appendix E which describes the data analysis methods in detail.

- **Prepare a Summary Sheet for Each Information Goal**

Summary sheets were an effective means of organizing and conveying the myriad of details associated with each of the twelve monitoring information goals.

- **State Procedures for Reviewing the Protocol**

Procedures were given for reviewing and updating the protocol. Recommendations provided by the DAP framework were closely followed.

- **Include Items to Clarify the Data Analysis Protocol**

Calculation procedures were clearly presented in the protocol and forms were provided for implementing statistical procedures.

CHAPTER VIII

SUMMARY AND CONCLUSIONS

SUMMARY

A few main themes of the thesis are briefly summarized here. Comprehensive summaries are placed throughout the thesis at the ends of sections or chapters.

The concept of a protocol is explored in Chapter II. Guidelines for protocol development are available in the literature, but they are widely scattered over a variety of disciplines. Although protocols for analyzing water quality data are beginning to emerge, there do not appear to be any general guidelines on what a water quality DAP should contain or how to write one.

Identification of information goals is the subject of Chapter III. It is emphasized that goals must be well defined if the rest of the DAP framework is to function properly.

Chapter IV addresses the complex issues related to handling of data record attributes such as serial correlation and seasonality. The current state of the art is presented and points of contention and uncertainty are brought out.

Choosing and interpreting statistical methods is discussed in Chapter V. It is pointed out that classical parametric statistical methods were originally developed for controlled experimental situations. The underlying assumptions of those methods frequently do not apply to water quality data.

The framework for development of groundwater quality DAPs is presented in Chapter VI. The framework provides clear and concise guidelines for protocol writers. For more detailed material, the protocol writer can refer to background information presented in Chapters III, IV and V.

Chapter VII is an overview of the case study. The framework was used to design a DAP for an IBM manufacturing facility. The bulk of the case study is presented as appendices.

CONCLUSIONS

A framework for the development of groundwater quality data analysis protocols is presented in this thesis. The framework describes the thought process that should be followed when writing a DAP. It allows people to write protocols in an informed manner. Background information for the framework is presented in an organized and readable fashion. Areas of debate are highlighted.

Practical application of the framework is demonstrated by using it to write a DAP for the IBM semiconductor manufacturing plant in East Fishkill, New York. Flexibility of the framework allowed the DAP to be site specific. Use of detailed background information (i.e., thesis chapters III, IV and V) added scientific validity to the protocol.

The IBM case study shows that data analysis protocols can be an effective means for ensuring that required information is obtained from monitoring programs. The IBM data analysis protocol focuses attention on information goals, and emphasizes the information content of data rather than data quantity.

RECOMMENDATIONS FOR FURTHER WORK

1. Improve the framework that is presented here by:
   - Refining it when additional information on the statistical analysis of water quality data becomes available.
   - Conducting more case studies.
   - Coding it as a computerized expert system.

2. Develop frameworks for writing protocols that address:
   - Analysis of existing data.
   - Analysis of poor quality data.
   - Information utilization.
   - Information presentation.

3. Expand research on these topics:
   - The validity of hypothesis tests.
   - The validity of sampling quarterly to avoid serial correlation.
   - The relationship between serial correlation and time scale.
   - Estimation of short-term *versus* long-term parameters.
   - Data visualization and spatial relationships.

4. Explore additional statistical data analysis methods such as equivalence tests.

CONCLUDING REMARKS

It is hoped that the framework presented in this thesis will elicit discussions among water quality professionals regarding standardization and use of statistical procedures. The framework is intended to be a sound starting point rather than the final word on how to write groundwater quality data analysis protocols.

REFERENCES

Aldrich, J.E. 1988. Development of a hospital radiation emergency protocol. Journal of the Canadian Association of Radiologists. 39:137-139.

Alexander, R.B. and R.A. Smith. 1988. Trends in lead concentrations in major U.S. rivers and their relation to historical changes in gasoline-lead consumption. Water Resources Bulletin. 24(3):557-569.

Anderson, S. and W.W. Hauck. 1983. A new procedure for testing equivalence in comparative bioavailability and other clinical trials. Communication in Statistics--Theory and Methods. 12(23):2663-2692.

Armstrong, J.W. and D.S. Becker. 1986. Protocol standardization for Puget Sound, Washington. Estuarine and coastal management--tools of the trade. Proceedings of the Tenth National Conference of the Coastal Society. October 12-15, 1986. New Orleans, LA.

ASTM. 1990. Standard practice for intralaboratory quality control procedures and a discussion on reporting low-level data. Designation: D 4210-89. 1990 Annual Book of Standards, Water and Environmental Technology. V. 11.01. American Society for Testing and Materials, Philadelphia, PA.

Barton, C., E. F. Vowinkel and J. P. Nawyn. 1987. Preliminary assessment of water quality and its relation to hydrogeology and land use: Potamac-Raritan-Magothy aquifer system, New Jersey. U.S. Geological Survey Water-Resources Investigations Report 87-4023.

Bayley, G.V. and J.M. Hammersley. 1946. The effective number of independent observations in an autocorrelated time series. Journal of the Royal Statistical Society. 8(1B):184-197.

Becker, D.S. and J.W. Armstrong. 1988. Development of regionally standardized protocols for marine environmental studies. Marine Pollution Bulletin. 19(7):310-313.

Becketti, S. and W. Gould. 1987. Rangefinder box plots. The American Statistician. 41(2):149.

Benjamini, Y. 1988. Opening the box of a boxplot. The American Statistician. 42(4):257-262.

Berryman, D., B. Bobee, D. Cluis and J. Haemmerli. 1988. Nonparametric test for trend detection in water quality time series. Water Resources Bulletin. 24(3):545-556.

Berthouex, P.M. and W.G. Hunter. 1975. Treatment plant monitoring programs: a preliminary analysis. Journal Water Pollution Control Federation. 47(8):2143-2156.

Berthouex, P.M., W.G. Hunter and L. Pallesen. 1987. Monitoring sewage treatment plants: some quality control aspects. Journal of Quality Technology. 10(4):139-149.

Berthouex, P.M. and I. Hau. 1991. Difficulties related to using extreme percentiles for water quality regulations. Research Journal of the Water Pollution Control Federation. 63(6):873-879.

Box, G.P. and G.M. Jenkins. 1976. Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco, CA.

Bradley, J.V. 1968. Distribution-Free Statistical Tests. Prentice-Hall, Englewood Cliffs, NJ.

Brinkley, K. 1987. Pronghorn hand-rearing protocol. Animal Keepers' Forum. 14(8):234-237.

Brown, D.A. 1990. EPA's resolution of the conflict between cleanup costs and the law in setting cleanup standards under Superfund. Columbia Journal of Environmental Law. 15(2):241-306.

Brown, P. and E.J. Mikkelsen. 1990. No Safe Place--Toxic Waste, Leukemia, and Community Action. University of California Press, Berkley, CA.

Canter, L.W. and R.C. Knox. 1986. Ground Water Pollution Control. Lewis Publishers, Chelsea, MI.

CDOH. 1992. Colorado water quality control act. Revised and reproduced by the Colorado Department of Health. November, 1990, updated August, 1992.

CFR. 1991. Code of Federal Regulations--Protection of the Environment, 40 parts 100 to 149, revised as of July 1, 1991. Published by the Office of the Federal Register National Archives and Records Administration as a special edition of the Federal Register.

Close, M.E. 1987. Effects of irrigation on water quality of a shallow unconfined aquifer. Water Resources Bulletin. 23(5):793-802.

Close, M.E. 1989. Effect of serial correlation on ground water quality sampling frequency. Water Resources Bulletin. 25(3):507-515.

Conover, W.J. and R.L. Iman. 1976. On some alternative procedures using ranks for the analysis of experimental designs. Communication in Statistics--Theory and Methods. A5(14):1349-1368.

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

Cooke, S.M., B.S. Gentry, S.H. Schroeder and D.S. Berry, eds. 1987a (yearly supplements through April, 1990). The Law of Hazardous Waste: Management, Cleanup, Liability, and Litigation. V. 1 of 3. Matthew Bender, New York, NY.

Cooke, S.M., B.S. Gentry, S.H. Schroeder and D.S. Berry, eds. 1987b (yearly supplements through April, 1990). The Law of Hazardous Waste: Management, Cleanup, Liability, and Litigation. V. 2 of 3. Matthew Bender, New York, NY.

Cotruvo J.A., F.A. Bell, Jr., G. Ruggiero and P. Tobin. 1982. Protocol development for potable reuse criteria and standards: summary of EPA workshop. Report of Workshop Proceedings on Protocol Development; Criteria and Standards for Potable Reuse and Feasible Alternatives. SCS Eng., Reston, VA.

Cowell, L.L. 1989. Development and validation of a ramping treadmill protocol for the on-line measurement of four aerobic parameters. Ph.D. Thesis, University of North Texas.

Currie, L.A. 1968. Limits for qualitative detection and quantitative determination--application to radiochemistry. Analytical Chemistry. 40(3):586-593.

Currie, L.A. 1988. Detection: overview of historical, societal, and technical issues. Detection in Analytical Chemistry--Importance, Theory, and Practice. L.A. Currie ed. American Chemical Society, Washington, D.C. pp. 1-62.

Daniel, W.W. 1991. Biostatistics, a foundation for analysis in the health sciences. 5th Ed. John Wiley and Sons, New York, NY.

Davis, J.C. 1977. Standardization and protocols of bioassays--their role and significance for monitoring, research and regulatory usage. Proceedings 3rd Aquatic Toxicity Workshop, Halifax N.S. Nov 2-3, 1976. Environmental Protection Service Technical Report No. EPS-5-AR-77-1. Halifax, Canada. pp. 1-14.

Dietz, E.J. and T.J. Killeen. 1981. A nonparametric multivariate test for monotone trend with pharmaceutical applications. Journal of the American Statistical Association. 76:169-174.

Dietz, E.J. 1989. Teaching regression in a nonparametrics statistics course. The American Statistician. 43(1):35-40.

du Toit, S.H.C., A.G.W. Steyn and R.H. Stumpf. 1986. Graphical Exploratory Data Analysis. Springer-Verlag, New York, NY.

Dubrovsky, N.M. and S.J. Deverel. 1989. Selenium in ground water of the central part of the western valley. Preliminary assessment of sources, distribution, and mobility of selenium in the San Joaquin Valley, California. U.S. Geological Survey Water-Resources Investigations Report 88-4186. pp. 35-52.

Duke, K.M. and R.G. Merrill, Jr. 1981. Development of new bioassay protocols. Management of Toxic Substances in Our Ecosystems--Taming the Medusa. B.W. Cornaby, ed. Ann Arbor Science Publishers.

Ellington, J.J., F.E. Stancil, Jr., W.D. Payne and C.D. Trusty. 1988. Interim protocol for measuring hydrolysis rate constants in aqueous solutions. U.S. Environmental Protection Agency Report No. EPA/600/3-88/014.

El-Shaarawi, A.H. and E. Damsleth. 1988. Parametric and nonparametric tests for dependent data. Water Resources Bulletin. 24(3):513-519.

El-Shaarawi, A.H. 1989. Inferences about the mean from censored water quality data. Water Resources Research. 25(4):685-690.

Fox J.P., T.R. Mongan and W.J. Miller. 1990. Trends in freshwater inflow to San Francisco Bay from the Sacramento-San Joaquin Delta. Water Resources Bulletin. 26(1):101-116.

Frigge, M., D.C. Hoaglin and B. Iglewicz. 1989. Some implementations of the boxplot. The American Statistician. 43(1):50-54.

Gasith, A., K.M. Jop, K.L. Dickson, T.F. Parkerton and S.A. Kaczmarek. 1988. Protocol for the identification of toxic fractions in industrial wastewater effluents. Aquatic Toxicology and Hazard Assessment: 10th Volume, ASTM STP 971. W.J. Adams, G.A. Chapman and W.G. Landis, eds. American Society for Testing and Materials, Philadelphia, PA. pp. 204-215.

Gibbons, J.D. 1985. Nonparametric Methods for Quantitative Analysis. 2nd Ed. American Sciences Press, Columbus, OH.

Gibbons, R.D. 1987a. Statistical prediction intervals for the evaluation of ground-water quality. Ground Water. 25(4):455-465.

Gibbons, R.D. 1987b. Statistical models for the analysis of volatile organic compounds in waste disposal sites. Ground Water. 25(5):572-580.

Gibbons, R.D. 1990. A general statistical procedure for ground-water detection monitoring at waste disposal facilities. Ground Water. 28(2):235-243.

Gibbons, R.D. 1991a. Statistical tolerance limits for ground-water monitoring. Ground Water. 29(4):1991.

Gibbons, R.D. 1991b. Some additional nonparametric prediction limits for ground-water detection monitoring at waste disposal facilities. Ground Water. 29(5):729-736.

Gibbons, R.D. and J. Baker. 1991. The properties of various statistical prediction intervals for ground-water detection monitoring. Journal of the Environmental Science and Health. A26(4):535-553.

Gilbert, R.O. 1987. Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold Company, New York, NY.

Gilliom, R.J., R.M. Hirsch and E.J. Gilroy. 1984. Effect of censoring trace-level water-quality data on trend-detection capability. Environmental Science and Technology. 18(7):530-535.

Gilliom, R.J. and D.R. Helsel. 1986. Estimation of distributional parameters for censored trace level water quality data; 1. Estimation techniques Water Resources Research. 22(2):135-146.

Glaser, J.A., D.L. Forest, G.D. McKee, S.A. Quave and W.L. Budde. 1981. Trace analyses for wastewaters. Environmental Science and Technology. 15(12):1426-1435.

Glass, E.A. 1988. Superfund and SARA: Are There Any Defenses Left? The Harvard Environmental Law Review. 12(2):385-464.

Greene, J.C., C.L. Bartels, W.J. Warren-Hicks, B.R. Parkhurst, G.L. Linder, S.A. Peterson and W.E. Miller. 1988. Protocols for short term toxicity screening of hazardous waste sites. U.S. Environmental Protection Agency Report No. EPA/600/3-88/029.

Hall, R.M., Jr., T. Watson, J.J. Davidson, D.R. Case and N.S. Bryson (Crowell and Moring). 1987. RCRA Hazardous Waste Handbook. 7th Ed. Government Institutes, Inc., Rockville, MD.

Hall, R.M., Jr., T. Watson, R.E. Schwartz, N.S. Bryson and R.C. Davis, Jr. (Crowell and Moring). 1990. Superfund Manual. 3rd Ed. Government Institute, Inc., Rockville, MD.

Harcum, J.B. 1990. Water-quality data analysis protocol development. Unpublished Ph.D. Thesis, Agricultural and Chemical Engineering Department, Colorado State University, Fort Collins, CO.

Harris, J., J.C. Loftis and R.H. Montgomery. 1987. Statistical methods for characterizing ground-water quality. Ground Water. 25(2):185-193.

Harris, J. 1988. Statistical analysis of ground water quality-interaction of deterministic and stochastic components. Unpublished Ph.D. Thesis, Agricultural and Chemical Engineering Department, Colorado State University, Fort Collins, CO.

Hedberg, A.M., N. Garcia, I.J. Trejus, S. Weinmann-Winkler, M.L. Gabriel and A.L. Lutz. 1988. Nutritional risk screening: development of a standardized protocol using dietetic technicians. Journal of the American Dietetic Association. 88(12):1553-1556.

Helsel, D.R. 1987. Advantages of nonparametric procedures for analysis of water quality data. Hydrological Sciences--Journal. 32(2):179-190.

Helsel, D.R. and R.M. Hirsch. 1988. Discussion: "Applicability of the t-test for detecting trends in water quality variables" by Robert H. Montgomery and Jim C. Loftis. Water Resources Bulletin. 24(1):201-204.

Helsel, D.R. 1990. Less than obvious--statistical treatment of data below the detection limit. Environmental Science and Technology. 24(12):1766-1774.

Helsel, D.R. and R.M. Hirsch. 1992. Statistical Methods in Water Resources. Elsevier Publishers, (not yet published).

Hem, J. 1985. Study and interpretation of the chemical characteristics of natural water. 3rd Ed. U.S. Geological Survey Water-Supply Paper 2254.

Hirsch, R.M. and E.J. Gilroy. 1985. Detectability of step trends in the rate of atmospheric deposition of sulfate. Water Resources Bulletin. 21(5):773-784.

Hirsch, R.M., R.B. Alexander and R.A. Smith. 1991. Selection of methods for the detection and estimation of trends in water quality. Water Resources Research. 27(5):803-813.

Hirsch, R.M., J.R. Slack and R.A. Smith. 1982. Techniques of trend analysis for monthly water quality data. Water Resources Research. 18(1):107-121.

Hirsch, R.M. and J.R. Slack. 1984. A nonparametric trend test for seasonal data with serial dependence. Water Resouces Research. 20(6):727-732.

Hirsch, R.M. 1988. Statistical methods and sampling design for estimating step trends in surface-water quality. Water Resources Bulletin. 24(3):493-503.

Hockman, K.K. and J.M. Lucas. 1987. Variability reduction through subvessel CUSUM control. Journal of Quality Technology. 19(3):113-121.

Hodges, J.L. and E.L. Lehmann. 1963. Estimates of location based on rank tests. The Annals of Mathematical Statistics. 34:598-611.

Hollander, M. and D.A. Wolfe. 1973. Nonparametric Statistical Methods. John Wiley and Sons, New York, NY.

Howell, G., and A. H. El-Shaarawi. 1991. An overview of the acidification of lakes in Atlantic Canada. Environmental Monitoring and Assessment. 17:323-338.

Hughes, J.P. and S.P. Millard. 1988. A tau-like test for trend in the presence of multiple censoring points. Water Resources Bulletin. 24(3):521-531.

Hunter, W.G. 1981. The practice of statistics: the real world is an idea whose time has come. The American Statistician. 35(2).

Iman, R.L. and W.J. Conover. 1983. A Modern Approach to Statistics. John Wiley and Sons, New York, NY.

Jones, D. and N. Matloff. 1986. Statistical hypothesis testing in biology; a contradiction in terms. FORUM: Journal of Economic Entomology. 79(5):1156-1160.

Jorgensen, E.P., ed., (Sierra Club Legal Defense Fund). 1989. The Poisoned Well. Island Press, Washington D.C.

Keith, L.H., W. Crummett, J. Deegan, Jr., R.A. Libby, J.K. Taylor and G. Wentler. 1983. Principles of environmental analysis. Analytical Chemistry. 55(14):2210-2218.

Keith, L.H. 1991. Environmental Sampling and Analysis: A Practical Guide. Lewis Publishers, Chelsea, MI.

Keller-McNulty, S. and M. McNulty. 1987. The independent pairs assumption in hypothesis tests based on rank correlation coefficients. The American Statistician. 41(1):40-41.

Kinnamon, D.E. 1992. Evaluation of offsite groundwater migration at a hazardous waste landfill in eastern Colorado. Unpublished M.S. Thesis, Department of Environmental Health, Colorado State University, Fort Collins, CO.

Kirchmer, C.J. 1988. Estimation of detection limits for environmental analytical procedures--a tutorial. Detection in Analytical Chemistry--Importance, Theory, and Practice. L.A. Currie ed. American Chemical Society, Washington, D.C.

Lambert, D., B. Peterson and I. Terpenning. 1991. Nondetects, detection limits, and the probability of detection. Journal of the American Statistical Association. 86(414):266-277.

Lanfear, K.J. and R.B. Alexander. 1990. Methodology to derive water-quality trends for use by the National Water Summary Program of the U.S. Geological Survey. U.S. Geological Survey Open-File Report 90-359, Reston, VA.

Lettenmaier, D.P. 1976. Detection of trends in water quality data from records with dependent observations. Water Resources Research. 12(5):1037-1046.

Lettenmaier, D.P. 1988. Multivariate nonparametric tests for trend in water quality. Water Resources Bulletin. 24(3):505-512.

Lettenmaier, D.P., E.R. Hooper, C. Wagoner and K.B. Faris. 1991. Trends in stream quality in the continental United States. Water Resources Research. 27(3):327-339.

Lipson, S.F. and P.T. Ellison. 1987. Development of protocols for collection of saliva samples under field conditions. (abstract only). American Journal of Physical Anthropology. 72(2):226.

Loftis, J.C., P.S. Porter and G. Settembre. 1987. Statistical analysis of industrial wastewater monitoring data. Journal of the Water Pollution Control Federation. 59(3):145-151.

Loftis, J.C. and C.H. Taylor. 1989. Testing for trend in water quality data. Proceedings International Symposium on the Design of Water Quality Information Systems, held in Fort Collins, CO on June 7-9, 1989. R.C. Ward, J.C. Loftis and G.B. McBride, eds. pp. 461-467. Information Series No. 61, Colorado Water Resources Research Institute, Fort Collins, CO.

Loftis, J.C., R.C. Ward, R.D. Phillips and C.H. Taylor. 1989. An evaluation of trend detection techniques for use in water quality monitoring programs. U.S. Environmental Protection Agency Report No. EPA/600/3-89/037.

Loftis, J.C., G.B. McBride and J.C. Ellis. 1991a. Considerations of scale in water quality monitoring and data analysis. Water Resources Bulletin. 27(2):1-10.

Loftis, J.C., C.H. Taylor, A.D. Newell and P.L. Chapman. 1991b. Multivariate trend testing of lake water quality. Water Resources Bulletin. 27(3):461-473.

Loftis, J.C., C.H. Taylor and P.L. Chapman. 1991c. Multivariate tests for trend in water quality. Water Resources Research. 27(7):1419-1429.

Lucas, J.M., 1982. Combined Shewhart-CUSUM quality control schemes. Journal of Quality Technology. 14(2):51-59.

Marmorek D.R., D.P. Bernard, C.H.R. Wedeles, G. Sutherland, J.A. Malanchuk and W.E. Fallon. 1989. A protocol for determining lake acidification pathways. Water, Air, and Soil Pollution. 44:235-257.

Marriott, F.H.C. 1990. A Dictionary of Statistical Terms. Longman Scientific and Technical, Essex, England, and John Wiley and Sons, Inc., New York, NY.

Maxwell R.J. and W.N. Marmer. 1983. Systematic protocol for the accumulation of fatty acid data from multiple tissue samples: tissue handling, lipid extraction and class separation, and capillary and chromatographic analysis. Lipids. 18(7).

McBride, G. and J. Loftis. 1991. Statistical Methods for Water Quality Sampling Programmes--notes for the short course. February 25-28 and March 11-23, 1991. Organized by Water Quality Centre, DSIR, Hamilton, New Zealand.

McBride, G.B., J.C. Loftis and N.C. Adkins. 1992. What do significance tests really tell us about the environment? Submitted to Environmental Management.

Mcleod, A.I., K.W. Hipel and F. Comancho. 1983. Trend assessment of water quality time series. Water Resources Bulletin. 19(4):537-547.

McNichols, R.J. and C.B. Davis. 1988. Statistical issues and problems in ground water detection monitoring at hazardous waste facilities. Ground Water Monitoring Review. Fall, 1988.

Mee, R.W. 1990. Confidence intervals for probabilities and tolerance regions based on a generalization of the Mann-Whitney statistic. Journal of the American Statistical Association. 85(411):793-800.

Mielke, P.W. and K.J. Berry. 1976. Multi-response permutation procedures for a priori classification. Communications in Statistics--Theory and Methods. A5(14):1409-1424.

Millard, S.P. and S.J. Deverel. 1988. Nonparametric statistical methods for comparing two sites based on data with multiple detection limits. Water Resources Research. 24(12):2087-2098.

Misuriello, H.P. 1987. A uniform procedure for the development and dissemination of building monitoring protocols. American Society of Heating, Refrigeration, and Air Conditioning Engineers--Transactions. 93(1):1619-1629.

Montgomery, R.H. and K.H. Reckhow. 1984. Techniques for detecting trends in lake water quality. Water Resources Bulletin. 20(1):43-52.

Montgomery, R.H. and J.C. Loftis. 1987. Applicability of the t-test for detecting trends in water quality variables. Water Resources Bulletin. 23(4):653-662.

Montgomery, R.H., J.C. Loftis and J. Harris. 1987. Statistical characteristics of ground-water quality variables. Ground Water. 25(2):176-184.

Montgomery, R.H. and J.C. Loftis. 1988. Reply to discussion by Dennis R. Helsel and Robert M. Hirsch "Applicability of the t-test for detection trends in water quality variables." Water Resources Bulletin. 24(1):205-207.

Natrella, M.G. 1960. The relation between confidence intervals and tests of significance. Statistical Issues--A Reader for the Behavioral Sciences. R.E. Kirk, ed. Brooks/Cole Publishing Company, Monterey, CA, 1972. pp. 113-117.

Nelson, D.W. and R.H. Dowdy (co-chairs of workshop and editors). 1988. Methods for ground water quality studies. Proceedings of a national workshop held at Arlington, Virginia on November 1-3, 1988. Published by Agricultural Research Division, University of Nebraska-Lincoln, Lincoln, NE.

Newell, C.J., L.P. Hopkins and P.B. Bedient. 1990. A hydrogeologic database for ground-water modeling. Ground Water. 28(5):703-714.

Newman, M.C., P.M. Dixon, B.B. Looney and J.E. Pinder III. 1989. Estimating mean and variance for environmental samples with below detection limit observations. Water Resouces Bulletin. 25(4):905-916.

NRC. 1990. Managing troubled waters--the role of marine environmental monitoring. A report by the Committee on a Systems Assessment of Marine Environmental Monitoring, Marine Board, Commission on Engineering and Technical Systems, National Research Council. National Academy Press, Washington D.C.

O'Brien, T.C. and B.J. Shapiro. 1968. Statistical significance--what? Statistical Issues--A Reader for the Behavioral Sciences. Brooks/Cole Publishing Company, Monterey, CA, 1972.

Opton, B.M. 1983. Environmental regulation: applicant behavior as a factor in obtaining permits. Ph.D. Thesis published as Colorado Water Resources Research Institute Information Series No. 51. Colorado State University, Fort Collins, Colorado.

Ozonoff, D. and L.I. Boden. 1987. Truth and consequences: health agency responses to environmental health problems. Science, Technology and Human Values. 12(3&4):70-77.

Patel, H.I. and G.D. Gupta. 1984. A problem of equivalence in clinical trials. Biometrical Journal. 26(5):471-474.

Pavur, R. 1988. Type I error rates for multiple comparison procedures with dependent data. The American Statistician. 42(3):171-173.

Perry, J.N. 1986. Multiple-comparison procedures: a dissenting view. FORUM: Journal of Economic Entomology. 79(5):1149-1155.

Phillips, R.D, J.C. Loftis and H.P. Hotto. 1989. User's manual for WQSTAT II--a water quality statistics program. Colorado State University, Fort Collins, CO.

Porter, P.S. 1986. Statistical analysis of water quality data affected by the limits of detection. Unpublished Ph.D. Thesis, Department of Agricultural and Chemical Engineering, Colorado State University, Fort Collins, CO.

Porter, P.S., R.C. Ward and H.F. Bell. 1988. The detection limit. Environmental Science and Technology. 22(8):856-861.

Porter, P.S. and R.C. Ward. 1991. Estimating central tendency from uncensored trace level measurements. Water Resources Bulletin. 27(4):687-700.

Quenouille, M.H. 1949. Approximate tests of correlation in time series. Journal of the Royal Statistical Society series B. 11:68-84.

Reed, M. and V.J. Bierman, Jr. 1983. Proceedings of a Workshop for the Development of a Scientific Protocol for Ocean Dumpsite Designation. Held at the W. Alton Jones Campus of the University of Rhode Island, February 14-17, Kingston, Rhode Island. U.S. Environmental Protection Agency Report No. ASA/RI-83-001.

Richards, R.P. 1989. Evaluation of some approaches to estimating non-point pollutant loads for unmonitored areas. Water Resources Bulletin. 25(4):891-904.

Sanders, T.G., R.C. Ward, J.C. Loftis, T.D. Steele, D.D. Adrian and V. Yevjevich. 1983. Design of Networks for Monitoring Water Quality. Water Resources Publications, Littleton, CO.

Sara, M.N. and R. Gibbons. 1991. Organization and analysis of water quality data. Practical Handbook of Ground-Water Monitoring. D.M. Neilson ed. Lewis publishers, Chelsea, MI. pp. 541-588.

Scariano, S.M. and J.M. Davenport. 1987. The effects of violations of independence assumptions in the one-way ANOVA. The American Statistician. 41(2):123-129.

Schaub, S.A. and C.P. Gerba. 1988. Guide standard and protocol for testing microbiological water purifiers. Proceedings--Conference on Point-of-Use Treatment of Drinking Water. Cincinnati, Ohio, October 6-8, 1987. U.S. Environmental Protection Agency Report No. EPA/600/9-88/012.

Schilling, E.G. and P.R. Nelson. 1976. The effect of non-normality on the control limits of X charts. Journal of Quality Technology. 8(4):183-188.

Schmid, C.F. 1983. Statistical graphics--design principles and practices. John Wiley and Sons, New York, NY.

Schroder, L.J. and H.E. Taylor. 1980. Contamination studies and testing protocol for polyethylene bottles. U.S. Geological Survey Open File Report 80-165.

Schubel, J.R. 1987. Program accountability--how do we measure progress?--A scientific view. Estuarine and coastal management--tools of the trade. Proceedings of the Tenth National Conference of The Coastal Society. October 12-15, 1986. New Orleans, LA. Copyright by The Coastal Society, 1987.

Schuk, W.W. 1986. Evaluation of a dissolved oxygen field test protocol. Aeration systems--design, testing, operation, and control. W.C. Boyle, ed. Noyes Publications, Park Ridge, NJ.

Sen, P.K. and M.L. Puri. 1977. Asymptotically distribution-free aligned rank order tests for composite hypotheses for general multivariate linear models. Z. Wahrscheinlichkeitstheor. Verw. Geb. 39:187-195.

Smith, D.G. and G.B. McBride. 1990. New Zealand's national water quality monitoring network--design and first year's operation. Water Resources Bulletin. 26(5):767-775.

Snedecor, G.W. and Cochran, W.G. 1989. Statistical Methods. 8th Ed. Iowa State University Press, Ames, IA.

Starks, T.H. and G.T. Flatman. 1991. RCRA ground-water monitoring decision procedures viewed as quality control schemes. Environmental Monitoring and Assessment. 16:19-37.

Statgraphics. 1991. Reference manual--Statgraphics V.5. STSC, Rockville, MD.

Steen, W.C. 1988. Project summary--interim protocol for measuring microbial transformation rate constants for suspended bacterial populations in aquatic systems. U.S. Environmental Protection Agency Report No. EPA/600/S3-88/007.

Taylor, J.K. 1987. Quality Assurance of Chemical Measurements. Lewis Publishers, Chelsea, MI.

Thomas, R.L. 1987. A protocol for the selection of process-oriented remedial options to control in situ sediment contaminants. Hydrobiologia. 149:247-258.

Thompson, C.R., G. Kats, P. Dawson and D. Doyle. 1981. Project summary--development of a protocol for testing the effects of toxic substances on plants. U.S. Environmental Protection Agency Report No. EPA-600/S3-81-006.

Tondeur, Y., W.F. Beckert, S. Billets and R.K. Mitchum. 1989. Method 8290: an analytical protocol for the multimedia characterization of polychlorinated dibenzodioxins and dibenzofurans by high-resolution gas chromatography/high-resolution mass spectrometry. Chemosphere. 18:119-131.

Tufte, E.R. 1983. The Visual Display of Quantitative Information. Graphic Press, Cheshire, CT.

Tufte, E.R. 1990. Envisioning Information. Graphics Press, Cheshire, CT.

Tukey, J.W. 1977. Exploratory Data Analysis. Addison-Wesley Publishing Company, Reading, MA.

U.S. Department of Labor. 1967. The consumer price index: history and techniques. U.S. Department of Labor, Bureau of Labor Statistics, Bulletin No. 1517.

U.S. EPA. 1985. Resource Document for the Ground-Water Monitoring Strategy Workshop. Office of Ground-Water Protection, U.S. Environmental Protection Agency, Washington D.C.

U.S. EPA. 1986. RCRA ground water monitoring technical enforcement guidance document (final). Office of Waste Programs Enforcement, U.S. Environmental Protection Agency, Washington, D.C. OSWER Report No. 9950.1.

U.S. EPA. 1988. Guidance on Remedial Actions for Contaminated Ground Water at Superfund Sites. Office of Emergency and Remedial Response, U.S. Environmental Protection Agency, Washington D.C.

U.S. EPA. 1989. Statistical analysis of ground-water monitoring data at RCRA facilities--interim final guidance. Office of Solid Waste, Waste Management Division, U.S. Environmental Protection Agency, Washington, D.C.

U.S. EPA. 1990. The Nation's Hazardous Waste Management Program at a Crossroads: The RCRA Implementation Study. Office of Solid Waste and Emergency Response, U.S. Environmental Protection Agency, Washington D.C.

Vaughan, W.J. and C.S. Russell. 1983. Monitoring point sources of pollution: answers and more questions from statistical quality control. The American Statistician. 37(4):476-487.

Velleman, P.F. and Hoaglin, D.C. 1981. Applications, Basics, and Computing of Exploratory Data Analysis. Duxbury Press, Boston, MA.

Wainer, H. and D. Thissen. 1981. Graphical data analysis. Annual Review of Psychology. 32:191-241.

Walker, W.W. 1991. Water quality trends at inflows to Everglades National park. Water Resources Bulletin. 27(1):59-72.

Wallis, J.R. and P.E. O'Connell. 1972. Small sample estimation of ρ_1. Water Resources Research. 8(3):707-212.

Ward, R.C., J.C. Loftis, H.P. DeLong and H.F. Bell. 1988. Groundwater quality: a data analysis protocol. Journal Water Pollution Control Federation. 60(11):1938-1945.

Ward, R.C. and J.C. Loftis. 1989. Monitoring systems for water quality. Critical Reviews in Environmental Control. 19(2):101-118.

Ward, R.C., J.C. Loftis and G.B. McBride. 1990. Design of Water Quality Monitoring Systems. Van Nostrand Reinhold Company, NY.

Weinberg, S. and K. Goldberg. 1990. Statistics for the Behavioral Sciences. Cambridge University Press, Cambridge, England.

Wilson, A.L. 1973. The performance characteristics of analytical methods--III. Talanta. 20:725-732.

Wolbach, C.D., R.R. Whitney and U.B. Spannagel. 1984. Project summary--design and development of a hazardous waste reactivity testing protocol. U.S. Environmental Protection Agency Report No. EPA-600/S2-84-057.

Wurman, R.S. 1989. Information Anxiety. Doubleday, New York, NY.

Zahn, D.A. and Isenberg, D.J. 1983. Nonstatistical aspects of statistical consulting. The American Statistician. 37(4), November.

Zar, J.H. 1984. Biostatistical Analysis. 2nd Ed. Prentice-Hall, Englewood Cliffs, NJ.

APPENDIX I

INITIAL SUBMITTAL OF GROUNDWATER MONITORING PROPOSAL

APPENDIX I

INITIAL SUBMITTAL OF GROUNDWATER MONITORING PROPOSAL

The groundwater monitoring proposal presented here was submitted to NYSDEC on June 5, 1992. The proposal is a draft of material that will appear in the body of IBM's final 373 permit.

Negotiations between IBM and NYSDEC were not a major factor in this initial groundwater monitoring proposal. It was written by IBM as a basis for future discussion.

IBM East Fishkill Groundwater Monitoring Proposal
NYS 373 Permit
Revision 1 - June 26, 1992

(1) Corrective Action Monitoring Wells.

The Corrective Action Monitoring Wells are intended to meet the following objectives:

- Process Control for Treatment - These monitoring wells will be used to detect significant changes in water quality (new compounds detected and/or significant change in concentration) in order to insure the adequacy of the treatment facilities. IBM will establish process control limits (treatable compounds, treatable concentration range, design flow rate, and design mass loading) for each of the treatment facilities. IBM will notify the NYSDEC within 7 days of detection and confirmation of a significant change in input concentration OR chemical composition which jeopardizes the treatability of groundwater.

- Status of the Source and Plume - The objective of these monitoring wells will be to evaluate the status and extent of the contamination plume. IBM will use this data to construct maps showing an estimate of the limits of each of the contamination plumes. The limit of each plume is defined as an annual median concentration equal to 5 ug/L of one key compound or sum of key compounds. In addition, IBM will determine whether there has been a significant change in water quality for each well and its key compounds or combination of key compounds. A change will be determined by utilizing a nonparametric medians comparison between the current 2 years of data and the previous 2 years of monitoring data. This analysis will be provided in the Annual Report.

Any additional compounds detected and confirmed above the Groundwater Protection Standard will be reported in the semi-annual status reports.

- Effectiveness of Remediation - The objective of these wells is to evaluate the effectiveness of remediation. Representative wells within the source zone and the extraction wells will be monitored. Data from these wells will be used to provide the mass of contaminants removed for each area of remediation and to provide a qualitative assessment of the source area. This information will be provided in the Annual Report.

(2) Appendix 33 Monitoring Wells.

Appendix 33 analysis will be performed on representative wells within source areas. This is to insure that all chemicals of concern are identified and properly monitored. If new constituents are detected, the procedure in Module Condition E.17.(c) (iv) of the IBM Endicott plan will apply. The raw data and any necessary explanation will be provided in the next semi-annual status report.

(3) Point of Compliance Wells

The objective of these monitoring wells is to measure water quality down-gradient of an Area of Concern (AOC) where flow may pass across a site boundary. The data collected from these wells will be compared against the Groundwater Protection Standard (GWPS) as described in the "Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities - Interim Final Guidance", EPA.1989. This comparison will be reported in the Annual Report.

If a significant change in water quality is detected and upon resampling the result is confirmed at a Point of Compliance, IBM will notify NYS-DEC and will take immediate action to determine the reason for the significant change in water quality and if necessary, take corrective action.

(4) Detection Monitoring.

The objective of these monitoring wells is to detect contamination which may be leaving the facility and moving offsite OR to detect contamination which may be coming from offsite onto the facility.

- ONSITE -> OFFSITE : This is addressed under Point of Compliance monitoring.

- OFFSITE -> ONSITE : If a significant change in water quality is detected, the monitor well will be resampled. Any compounds detected and confirmed above the Groundwater Protection Standard will be addressed in the next semi-annual status report.

(5) Background Monitoring.

The objective of these monitor wells is to provide a qualitative basis of comparison with the wells monitored under Appendix 33 monitoring wells. These wells will be included in the annual Appendix 33 sampling for 3 years.

Results will be provided in the next semi-annual Report.

(6) Interim Monitoring Program Wells

The objective of these wells is to provide data for investigative studies associated with the NYS Part 373 Corrective Action Program. Water quality samples will be collected for organic and inorganic analysis to meet the needs of the investigation. All raw data collected and a project status will be reported in the next semi-annual status report.

(7) Hydraulic Effectiveness Monitoring Wells

All onsite (Main Site and West Complex) monitor wells will be measured quarterly to determine the hydraulic effectiveness of remediation.

IBM East Fishkill - Area A Remediation Area
Water Quality Monitoring

Well	Types of Monitoring: Corrective Action[1]	Point of Compliance	Appendix 33[2]	Detection Monitoring	Background Monitoring	Interim Monitoring Program	Sampling Frequency (#/year)	Key Compounds
702	S						2	Tetrachloroethene
007	S						2	Trichloroethene
998	S						2	cis-1,2-Dichloroethene
999	S						2	
720	S						2	
761	S						2	
099	S						2	
763	S						2	
204	S						2	
714	S		X				2	
Sump	S,P,E		X				12	
016	S,E						2	
103	S,E						2	
104	S,E						2	

1 p = Monitored for process control
s = Monitored to assess status of the source &/or plume
e = Monitored to assess the effectiveness of remediation

2 Annual (1x/yr)

IBM East Fishkill - Area A Remediation Area (cont)

Water Quality Monitoring

Well	Types of Monitoring: Corrective[1] Action	Point of Compliance	Appendix[2] 33	Detection Monitoring	Background Monitoring	Interim Monitoring Program	Sampling Frequency (#/year)	Key Compounds
744						X	4	
745						X	4	
015						X	4	
706	S		X				2	
SW						X	12	

IBM East Fishkill -Area B Remediation Area
Water Quality Monitoring

Well	Types of Monitoring					Interim	Sampling	Key
	Corrective[1] Action	Point of Compliance	Appendix[2] 33	Detection Monitoring	Background Monitoring	Monitoring Program	Frequency (#/year)	Compounds
863	S,P,E		X				12	Tetrachloroethene
834	S						4	
571	S						4	
064	S	X					4	
065	S	X					4	
863 GAC	P						12	

IBM East Fishkill -Area C Remediation Area (Landfill)
Water Quality Monitoring

Well	*Types of Monitoring* *Corrective[1] Action*	*Point of Compliance*	*Appendix[2] 33*	*Detection Monitoring*	*Background Monitoring*	*Interim Monitoring Program*	*Sampling Frequency (#/year)*	*Key Compounds*
563	S						2	Tetrachloroethene
561	S						2	Trichloroethene
559	S		X				2	cis-1,2-Dichloroethene
053	S						2	Vinyl Chloride
054	S						2	Freon TF
932	S						2	
944	S						2	
943	S						2	
069	S						2	

IBM East Fishkill -Area C B/330 Investigative Area
Water Quality Monitoring

Well	Types of Monitoring					Interim	Sampling	Key
	Corrective[1] Action	Point of Compliance	Appendix[2] 33	Detection Monitoring	Background Monitoring	Monitoring Program	Frequency (#/year)	Compounds
572						X	1	Tetrachloroethene
181			X			X	4	Trichloroethene
756						X	4	cis-1,2 Dichloroethene
062						X	4	Vinyl Chloride
063						X	4	Freon TF
061						X	4	
059						X	4	
060						X	4	
143						X	2	
145						X	2	
146						X	2	
929						X	2	
930						X	2	
931						X	2	
953						X	2	

IBM East Fishkill -Area C B/330 Investigative Area
Water Quality Monitoring

Well	Types of Monitoring: Corrective[1] Action	Point of Compliance	Appendix[2] 33	Detection Monitoring	Background Monitoring	Interim Monitoring Program	Sampling Frequency (#/year)	Key Compounds
574						X	1	Tetrachloroethene
178						X	4	Trichloroethene
576						X	4	cis-1,2 Dichloroethene
921						X	2	Vinyl Chloride
922						X	2	Freon TF
923						X	2	
924						X	2	
954						X	2	
955						X	2	
956						X	2	
957						X	2	
965						X	2	
966						X	2	
941						X	2	
PW -7	S,P,E		X				12	

IBM East Fishkill -Area D Remediation Area
Water Quality Monitoring

	Types of Monitoring							
Well	**Corrective[1] Action**	**Point of Compliance**	**Appendix[2] 33**	**Detection Monitoring**	**Background Monitoring**	**Interim Monitoring Program**	**Sampling Frequency (#/year)**	**Key Compounds**
032	P,S,E		X				12	Tetrachloroethene
757	S						4	Trichloroethene
067	S	X					4	cis-1,2-Dichloroethene
066	S	X					4	

IBM East Fishkill -SEQ Remediation Area
Water Quality Monitoring

Well	Types of Monitoring					Interim	Sampling	Key
	Corrective[1] Action	Point of Compliance	Appendix[2] 33	Detection Monitoring	Background Monitoring	Monitoring Program	Frequency (#/year)	Compounds
048	S						4	Tetrachloroethene
552	S		X				4	Trichloroethene
543	S						4	cis-1,2-Dichloroethene
545	S						4	
990	S	X					12	
987	S	X					12	
520	S						4	
502	S						4	
503	S						4	
859						X	4	
860						X	4	
861						X	4	
862						X	4	

IBM East Fishkill - Bedrock Aquifer
Water Quality Monitoring

Well	*Types of Monitoring* Corrective[1] Action	*Point of Compliance*	*Appendix[2] 33*	*Detection Monitoring*	*Background Monitoring*	*Interim Monitoring Program*	*Sampling Frequency (#/year)*	*Key Compounds*
PW-1	P,S,E						12	Tetrachloroethene
PW-2	P,S,E		X				12	Trichloroethene
PW-4	P,S,E		X				12	cis-1,2-Dichlorethene
PW-5/5A	P,S,E						12	Vinyl Chloride
PW-6	P,S						12	Freon TF
PW-7	P,S,E		X				12	
PW-9	P,S,E						when on	
GAC-Input	P		X				12	
113	S						4	
535	S						4	
958	S						4	
515	S						4	
963	S						4	
964	S						4	

IBM East Fishkill - Bedrock Aquifer (cont)
Water Quality Monitoring

Well	Corrective[1] Action	Point of Compliance	Appendix[2] 33	Detection Monitoring	Background Monitoring	Interim Monitoring Program	Sampling Frequency (#/year)	Key Compounds
	Types of Monitoring							
521	S						4	Tetrachloroethene
857	S						4	Trichloroethene
858	S						4	cis-1,2-Dichloroethene
166	S			X			2	Vinyl Chloride
167	S			X			2	Freon-TF
087	S			X			2	
085	S			X			2	
722	S						2	
711	S						2	
105	S						2	
106	S						2	
704	S						2	
108	S						2	
109	S						2	
959	S		X				2	

IBM East Fishkill - Bedrock Aquifer (cont)

Water Quality Monitoring

Well	*Corrective[1] Action*	*Point of Compliance*	*Appendix[2] 33*	*Detection Monitoring*	*Background Monitoring*	*Interim Monitoring Program*	*Sampling Frequency (#/year)*	*Key Compounds*
	Types of Monitoring							
965	S						2	Tetrachloroethene
966	S						2	Trichloroethene
564	S						2	cis-1,2-Dichloroethene
563	S						2	Vinyl Chloride
559	S						2	Freon TF
932	S						2	
944	S						2	
943	S						2	
145	S						2	
146	S						2	
929	S						2	
930	S						2	
931	S						2	
941	S		X				2	

IBM East Fishkill - Bedrock Aquifer (cont)
Water Quality Monitoring

Well	Types of Monitoring: Corrective[1] Action	Point of Compliance	Appendix[2] 33	Detection Monitoring	Background Monitoring	Interim Monitoring Program	Sampling Frequency (#/year)	Key Compounds
953	S						2	Tetrachloroethene
143	S						2	Trichloroethene
156	S						2	cis-1,2-Dichloroethene
945	S						2	Vinyl Chloride
952	S						2	Freon TF
737	S						2	
742	S						2	
173	S						2	
150	S			X			2	
779	S			X			2	
778	S			X			2	
777	S			X			2	
739	S						2	
716	S		X				2	

IBM East Fishkill - Building 322 Investigative Area
Water Quality Monitoring

Well	Types of Monitoring: Corrective[1] Action	Point of Compliance	Appendix[2] 33	Detection Monitoring	Background Monitoring	Interim Monitoring Program	Sampling Frequency (#/year)	Key Compounds
735						X	4	Freon TF
740						X	4	Freon 123a
766						X	4	Tetrachloroethene
767						X	4	Trichloroethene
768						X	4	cis-1,2-Dichloroethene
769						X	4	
770						X	4	
771						X	4	
772						X	4	
773						X	4	
754						X	4	
755						X	4	
792						X	4	
793						X	4	
781						X	4	
203						X	4	

IBM East Fishkill - Building 322 Investigative Area (cont.)
Water Quality Monitoring

Well	Types of Monitoring: Corrective[1] Action	Point of Compliance	Appendix[2] 33	Detection Monitoring	Background Monitoring	Interim Monitoring Program	Sampling Frequency (#/year)	Key Compounds
151						X	4	Freon TF
949						X	4	Freon 123a
743						X	4	Tetrachloroethene
161						X	4	Trichloroethene
035						X	4	cis-1,2-Dichloroethene
068						X	4	
774						X	4	
775						X	4	
791						X	4	
789						X	4	
202						X	4	

Chemical Constituents/Monitoring Parameters		
Constituent	**Method**[1]	**Conc. Limit (ug/L)**[2]
Benzene	8021	0.7
Chlorobenzene	8021	5.0
1,2-Dichlorobenzene	8021	5.0
1,2,3-Trichlorobenzene	8021	5.0
1,2,4-Trichlorobenzene	8021	5.0
1,3-Dichlorobenzene	8021	5.0
1,4-Dichlorobenzene	8021	5.0
Dichlorodifluoromethane	8021	5.0
1,1-Dichloroethene	8021	5.0
Methylene Chloride	8021	5.0
Ethylbenzene	8021	5.0
Vinyl Chloride	8021	2.0
1,2-Dichloro-1,2,2-trifluoroethane	8021	5.0
1,1,2,2-Tetrachloroethene	8021	5.0
Freon TF	8021	5.0
1,1,2-Trichloroethene	8021	5.0
cis-1,2-Dichloroethene	8021	5.0
Fluoride[3]	4500-F C.[4]	1500.0
Arsenic (total)[5]	7060	25.0
Chromium (total)[6]	7191	50.0
Zinc (total)[7]	6010	300.0

1 EPA Method SW-846.

2 Groundwater Protection Concentrations for Corrective Measures.

3 Area "A", Area "C" (Bldg. 330) & Building 322 investigative area.

4 APHA, AWWA, WPCF, American Public Health Association. *Standard Methods For the Examination of Water and Wastewater*. 17th ed., Washington, DC, p. 4-87, 1989.

5 Building 322 investigative area.

6 Area "A".

7 Area "C" (Bldg. 330) investigative area.

APPENDIX II

NEGOTIATIONS BETWEEN NYSDEC AND IBM REGARDING MODIFICATIONS TO THE INITIAL GROUNDWATER MONITORING PROPOSAL

APPENDIX II

NEGOTIATIONS BETWEEN NYSDEC AND IBM REGARDING MODIFICATIONS TO THE INITIAL GROUNDWATER MONITORING PROPOSAL

Several meetings were held between IBM and NYSDEC to discuss modifications to the initial submittal of the groundwater monitoring proposal. Those negotiations are summarized below:

- Nomenclature of wells was one of the main topics of discussion. In their initial submittal, IBM labeled several wells as "detection monitoring" wells. Although the name accurately describes the function of the wells, it conflicts with the regulatory definition of detection monitoring. To avoid future confusion over nomenclature, IBM agreed to rename the wells. On a similar note, IBM agreed to rename the "West Complex" wells as the "West Complex Detection Monitoring" wells, because they meet the regulatory definition of detection monitoring.

- Sampling frequencies were also a matter of debate. Frequencies proposed by IBM in the initial submittal ranged from monthly at the more important wells to yearly for Appendix 33 sampling. NYSDEC agreed that Appendix 33 sampling should occur annually as per regulations. They felt, however, that the semi-annual sampling frequencies proposed by IBM should be changed to quarterly. IBM explained their philosophy of choosing sampling frequencies that are consistent with regulatory requirements and meet information goals. (See section VI of the data analysis protocol for specifics of this approach). NYSDEC and IBM agreed that it is better to have a combination of monthly, quarterly and semi-annual frequencies that are based on factors such as risk and historical water quality, than to have across-the-board quarterly sampling.

- NYSDEC felt that a few wells should be added to the monitoring system. IBM agreed to add some of the wells.

- NYSDEC was undecided on whether information concerning wells in areas undergoing active investigation should be located in the body of the permit or in the appendices. Their final decision was that it should be placed in the body.

- NYSDEC wanted IBM to write a summary report of the areas under investigation. IBM agreed to write the report.

The spreadsheet shown on the following three pages was a used as a tool during the negotiations. The spreadsheet compares sampling frequencies and locations initially proposed by IBM to those preferred by NYSDEC.

Well	IBM CA STAT.	IBM CA EFF.	IBM CA PROC.	IBM Pt. of Compliance	IBM Detect. Monitor.	IBM Backgrd. Monitor.	IBM Appd33	IBM Interim Monitor.	IBM Frequency	DEC Frequency	DEC CAM CA STAT.	DEC CRM CA EFF.	DEC Process CA PROC.	DEC Pt. of Compliance	DEC Detect. Monitor.	DEC Backgrd. Monitor.	DEC Appd33
G863Area B			X						12	12			X				
GACIN/316			X				X		12	12			X				X
JJHS					X		X		4	2	X				X		X
P1	X	X	X						12	12	X	X	X				
P2	X	X	X				X		12	12	X	X	X				X
P4	X	X	X				X		12	12	X	X	X				X
P5	X	X	X						12	12	X	X	X				
P6	X		X						12	12	X		X				
P7	X	X	X				X		12	12	X	X	X				X
P9	X	X	X						WHEN ON	12	X	X	X				X
SUMP/B384	X	X	X				X		12	12	X	X	X				X
SW								X	12								
7	X								2	2	X						
15								X	4								
16	X	X							2	2	X	X					
32	X	X	X				X		12	12	X	X					X
35								X	4								
48										4	X			X			X
48	X								4	2	X						
53	X								2	2	X						
54	X								2	2	X						
59								X	4								
60								X	4								
61								X	4								
62								X	4								
63								X	4								
64	X			X					4	4	X			X			X
65	X			X					4	4	X			X			X
66	X			X					4	4	X			X			X
67	X			X					4	4	X			X			X
68								X	4								
69	X								2	2	X						
85	X				X				2	2	X				X		
87	X				X				2	2	X				X		
99	X								2	2	X						
103	X	X							2	2	X	X					
104	X	X							2	2	X	X					
105	X								2	2	X						
106	X								2	2	X						
108	X								2	2	X						
109	X								2	2	X						
113	X								4	2	X						
143	X							X	2	2	X						
145	X							X	2	2	X						
146	X							X	2	2	X						
150	X				X				2	2	X				X		
151								X	4								
156	X								2	2	X						
161								X	4								
166	X				X				2	2	X				X		
167	X				X				2	2	X				X		
172										4	X			X			X
173	X								2	2	X						
178								X	4								
181							X	X	4								X
202								X	4								
203								X	4								
204	X								2	2	X						
400					X				2	2					X		
401					X				2	2					X		
402					X				2	2					X		
403					X				2	2					X		
404					X				2	2					X		

Well	IBM CA STAT.	IBM CA EFF.	IBM CA PROC.	IBM Pt. of Compliance	IBM Detect. Monitor.	IBM Backgrd. Monitor.	IBM Appd33	IBM Interim Monitor.	IBM Frequency	DEC Frequency	DEC CAM CA STAT.	DEC CRM CA EFF.	DEC Process CA PROC.	DEC Pt. of Compliance	DEC Detect. Monitor.	DEC Backgrd. Monitor.	DEC Appd33
407						X			1	2						X	
408					X				2	2					X		
410					X				2	2					X		
412					X				2	2					X		
415					X				2	2					X		
416					X				2	2					X		
417					X				2	2					X		
423					X				2	2					X		
424					X		X		2	2	X				X		X
502	X								4	4	X			X			X
503	X								4	4	X			X			X
515	X								4	2	X						
520	X								4	2	X						
521	X								4	2	X						
535	X								4	2	X						
543	X								4	2	X						
545	X								4	2	X						
552	X						X		4	2	X						X
559	X						X		2	2	X						X
561	X								2	2	X						
563	X								2	2	X						
564	X								2	2	X						
571	X								4								
572								X	1								
574								X	1								
576								X	4								
701										2	X						
702	X								2	2	X						
704	X								2	2	X						
706	X						X		2	2	X						X
711	X								2	2	X						
714	X						X		2	2	X	X					X
716	X						X		2	2	X						X
720	X								2	2	X						
722	X								2	2	X						
735								X	4								
737	X								2	2	X						
739	X								2	2	X						
740								X	4								
742	X								2	2	X						
743								X	4								
744									4								
745									4								
751										2	X						
753										2	X						
754								X	4								
755								X	4								
756								X	4								
757	X								4	2	X						
761	X								2	2	X						
763	X								2	2	X						X
766								X	4								
767								X	4								
768								X	4								
769								X	4								
770								X	4								
771								X	4								
772								X	4								
773								X	4								
774								X	4								
775								X	4								
777	X				X				2	2	X				X		
778	X				X				2	2	X				X		

Well	IBM CA STAT.	IBM CA EFF.	IBM CA PROC.	IBM Pt. of Compliance	IBM Detect. Monitor.	IBM Backgrd. Monitor.	IBM Appd33	IBM Interim Monitor.	*IBM* Frequency	*DEC* Frequency	DEC *CAM* CA STAT.	DEC *CRM* CA EFF.	DEC *Process* CA PROC.	DEC Pt. of Compliance	DEC Detect. Monitor.	DEC Backgrd. Monitor.	DEC Appd33
779	X				X				2	2	X				X		
781								X	4								
789								X	4								
791								X	4								
792								X	4								
793								X	4								
834	X								4	2	X						
857	X								4	2	X						
858	X								4	4	X			X			X
859								X	4								
860								X	4								
861								X	4								
862								X	4								
863	X	X	X				X		12	12	X	X					
921								X	2								
922								X	2								
923								X	2								
924								X	2								
929	X							X	2	2	X						
930	X							X	2	2	X						
931	X							X	2	2	X						
932	X								2	2	X						
941	X						X	X	2	2	X						X
943	X								2	2	X						
944	X								2	2	X						
945	X								2	2	X						
949								X	4								
952	X								2	2	X						
953	X							X	2	2	X						
954								X	2								
955								X	2								
956								X	2								
957								X	2								
958	X								4	2	X						
959	X						X		2	2	X						X
963	X								4	2	X						
964	X								4	2	X						
965	X							X	2	2	X						
966	X							X	2	2	X						
987	X			X					12	4	X			X			X
990	X			X					12	4	X			X			X
998	X								2	2	X						
999	X								2	2	X						

APPENDIX III

FINAL SUBMITTAL OF GROUNDWATER MONITORING PROPOSAL

FOR NYS 373 PERMIT

APPENDIX III

FINAL SUBMITTAL OF GROUNDWATER MONITORING PROPOSAL FOR NYS 373 PERMIT

The final groundwater monitoring proposal had not yet been submitted at the time this thesis was published. The issue of sampling frequencies was still being debated. NYSDEC was unsure of IBM's proposed semi-annual sampling frequencies at 30 of the wells. At NYSDEC's request, IBM prepared a detailed report which examined the information content of various sampling frequencies. IBM evaluated changes in water quality using the signed rank test and visual inspection of time series plots. They concluded that the information content of semi-annual water quality data is quite similar to that of monthly or quarterly data at the wells they reviewed.

APPENDIX IV

DATA ANALYSIS PROTOCOL FOR GROUNDWATER MONITORING CONDUCTED IN RESPONSE TO NYS 373 PERMIT--IBM EAST FISHKILL FACILITY

TABLE OF CONTENTS

page APP-

I.	INTRODUCTION	35
	A. Purpose	35
	B. Scope	36
	C. Organization of Protocol	37
II.	INFORMATION GOALS	37
	A. General	37
	B. Regulatory Basis for the Groundwater Monitoring Program	38
	C. Summary of Information Goals	40
III.	DATA RECORD ATTRIBUTES	42
	A. General	42
	B. Serial Correlation	43
	C. Seasonality	43
	D. Nonnormality	44
	E. Censoring	44
	F. Data Record Attributes Resulting From Operational Difficulties	45
IV.	DATA ANALYSIS METHODS	46
	A. General	46
	B. Computational Methods	47
	C. Graphical Methods	47
	D. Direct Comparison with Fixed Limits	48
	E. Statistical Methods	50
	1. Control Charts	50
	2. Interval Estimation of the Difference in Medians	53

TABLE OF CONTENTS, cont.

3. Interval Estimation of the Median 54

V. CRITERIA FOR CHOOSING COMPOUNDS FOR LABORATORY ANALYSIS, DATA SUMMARIZATION AND DATA ANALYSIS 55

A. Routinely Monitored Compounds 55

B. Selected Compounds 55

C. Key Compounds 56

VI. CRITERIA FOR CHOOSING SAMPLING INTERVALS 56

VII. PROCEDURES FOR REVISING DATA ANALYSIS PROTOCOL 56

VIII. REFERENCES 58

APPENDIX A--DATA ANALYSIS PROTOCOL SUMMARY SHEETS 61

APPENDIX B--SAMPLING FREQUENCIES 80

Table B-1. Sampling frequencies (per year) --area A remediation area. 81

Table B-2. Sampling frequencies (per year) --area B remediation area. 83

Table B-3. Sampling frequencies (per year) --area C remediation area. 84

Table B-4. Sampling frequencies (per year) --area D remediation area. 85

Table B-5. Sampling frequencies (per year) --SEQ remediation area. 86

Table B-6. Sampling frequencies (per year) --area C B/330 investigative area. 87

Table B-7. Sampling frequencies (per year) --building 322 investigative area. 89

Table B-8. Sampling frequencies (per year) --bedrock aquifer. 91

Table B-9. Sampling frequencies (per year) --west complex. 95

TABLE OF CONTENTS, cont.

APPENDIX C--CATEGORIES OF COMPOUNDS 96

Table C-1. "Appendix 33" compounds. 97

Table C-2. Routinely monitored compounds. 98

Table C-3. Selected compounds. 100

Table C-4. Key compounds. 102

APPENDIX D--REPORTING FORMATS 103

Figure D-1. Reporting format for combined Shewhart-CUSUM control charts. 104

Figure D-2. Reporting format for Hodges-Lehmann estimates and confidence intervals. 105

Figure D-3. Reporting format for time series plots. 106

Figure D-4. Reporting format for interval estimates of the median. 107

APPENDIX E--DATA ANALYSIS METHODS 108

Appendix E-I. Combined Shewhart-CUSUM control charts. 109

Appendix E-II. The Hodges-Lehmann estimator and confidence interval. 120

Appendix E-III. Interval estimate of the median. 124

GLOSSARY OF ACRONYMS

AOC	area of concern
ASTM	American Society for Testing and Materials
CSU	Colorado State University
CUSUM	cumulative sum
DAP	data analysis protocol
GC/MS	gas chromatograph/mass spectrometer
GWPS	groundwater protection standard
MDL	method detection limit
NYS	New York State
NYSDEC	New York State Department of Environmental Conservation
RCRA	Resource Conservation and Recovery Act
SEQ	southeast quadrant

I. INTRODUCTION

A. Purpose

This document gives procedures for using data analysis methods to obtain information from IBM's groundwater monitoring program. Water quality data for the program will be obtained in response the NYSDEC part 373 permit requirements.

The data analysis protocol (DAP) will help to ensure that all collected data has a specified purpose, and that statistics is properly used both in a theoretical and a practical sense.

It is felt that the DAP will provide numerous benefits including:

1. Those involved in data collection will develop a better understanding of the importance of producing high quality data.
2. The subjectivity of statistical analysis will be attenuated because statistical methods are specified prior to data collection.
3. Economic benefits will be realized because only data that contributes to useful information will be collected.
4. Continuity of data analysis in the event of employee turnover will be facilitated.
5. Data analysis methods will be correctly applied thereby resulting in reliable information.

6. Because they are well documented, data analysis procedures will be auditable.
7. Individuals who are involved with management and operation of the monitoring program will have a basis for communication.
8. Water quality information will be extracted from the data as soon as possible.
9. Monitoring system operation will be driven by information goals rather than by politics or short term crises.

This protocol is a joint effort between Colorado State University (CSU) researchers and IBM groundwater staff. It represents IBM's philosophy to "meet or exceed all regulatory requirements." It also meets RCRA goals of being protective of human health and the environment.

B. Scope

The original intent of this protocol was to provide guidelines for using statistical methods to analyze groundwater quality data. Soon after development of the protocol began, it become apparent that the scope should be widened to include non-statistical methods. It was also evident that issues closely related to data analysis, such as sampling frequencies and reporting formats, should be included in the protocol.

The primary intended users of this DAP are groundwater professionals at the IBM East Fishkill Facility. The protocol will also be useful to general managers, field samplers and laboratory personnel.

This DAP primarily addresses monitoring from a long-term management point of view. DAPs for investigative studies will be written separately.

C. Organization of Protocol

A large portion of this protocol is presented in appendices. Appendix A consists of DAP summary sheets which are the supporting structure of the rest of the protocol. Protocol organization is clearly shown in the table of contents.

II. INFORMATION GOALS

A. General

The importance of information goal identification for water quality monitoring programs has been noted by many authors. For example:

> "... it is simply a waste of money to monitor without a clear relationship between the information to be produced and its use within the management agency's decision making process" (Ward *et al.*, 1990).

"Many monitoring programs are ineffective because they devote too little attention to the formulation of clear goals and objectives..." (NRC, 1990).

Identification of information goals for the IBM East Fishkill groundwater monitoring program was a challenging process. The applicable environmental regulations are not written in specific enough terms to be used directly as information goals. Therefore, it was left up to regulators from the NYSDEC and representatives from IBM to decide what information is needed from the monitoring program and what the monitoring system can produce.

Discussions between regulators and IBM representatives proved to be essential to formulation of information goals. As stated by Wurman (1989), "Conversations are organic: their very structure is a give-and-take that allows understanding to happen."

B. Regulatory Basis for the Groundwater Monitoring Program

Laws and regulations were carefully reviewed even though information goals could not be derived directly from them. The review was helpful because it provided awareness of legislative intent. Also, knowledge of the laws and regulations proved useful during conversations with NYSDEC regulators.

The purpose of the following summary is to guide the reader through the maze of laws and regulations that are applicable for purposes of this data analysis protocol.

U.S. Code

Title 42: The Public Health and Welfare

Chapter 82: Solid Waste Disposal

Subchapter III: Hazardous Waste Management

Section 6924: Standards applicable to owners and operators of hazardous waste treatment, storage, and disposal facilities

(p) Ground water monitoring

Section 6926: Authorized State hazardous waste programs

(b) Authorization of State programs

The State of New York was given authorization to manage their own hazardous waste program: 51 Federal Register 17737 (May 15, 1986).

Code of Federal Regulations

Title 40: Protection of Environment

Chapter 1: Environmental Protection Agency

Subchapter I: Solid Wastes

Part 264: Standards for owners and operators of hazardous waste treatment, storage, and disposal facilities

Part 271: Requirements for authorization of State hazardous waste programs

---

New York Environmental Conservation Law

Article 27: Waste and Refuse

Title 9: Industrial Hazardous Waste Management

Section 27-0911: Standards applicable to owners and operators of hazardous waste treatment, storage, and disposal facilities

---

New York Code of Rules and Regulations

Title 6: Environmental Conservation

Subpart 373-2: Final status standards for owners and operators of hazardous waste treatment, storage, and disposal facilities

Section 373-2.6: Releases from solid waste management units

(1) Corrective action for solid waste management units

New York Permit Writers' Guide

373 Generic (5/14/91)

Module III: Corrective Action Requirements for Solid Waste Management Units and Areas of Concern

E.: Corrective Action Requirements

C. Summary of Information Goals

Information goals that were identified for this data analysis protocol are listed here. Summary sheets for each information goal are presented in Appendix A.

INFORMATION GOAL #1: To ensure adequacy of the treatment facility by screening for the presence of new compounds.

INFORMATION GOAL #2: To ensure adequacy of the treatment facility by detecting changes in input concentrations.

INFORMATION GOAL #3: To determine the lateral extent of contamination plumes.

INFORMATION GOAL #4: To determine the mass of contaminants removed at each remediation area.

INFORMATION GOAL #5: To characterize contaminant concentrations over time within the plumes.

INFORMATION GOAL #6: To determine the hydraulic effectiveness of remediation.

INFORMATION GOAL #7: To measure water quality down-gradient of AOCs where flow may pass across a site boundary.

INFORMATION GOAL #8: To confirm previous monitoring results that indicate that contaminants are not migrating offsite from onsite.

INFORMATION GOAL #9: To confirm previous monitoring results that indicate that contaminants are not migrating onsite from offsite.

INFORMATION GOAL #10: To ensure that all chemicals of concern are properly identified and monitored.

INFORMATION GOAL #11: To provide a qualitative basis for comparison with data from "Appendix 33" monitoring of contaminated wells.

INFORMATION GOAL #12: To provide data for investigative studies associated with the NYS Part 373 Corrective Action Program.

III. DATA RECORD ATTRIBUTES

A. General

Data record attributes are characteristics of data that can complicate statistical analysis (Bell and Delong, 1988). They are the result of three factors: (1) statistical characteristics, (2) technological limitations and (3) operational difficulties.

Three statistical characteristics are discussed in this section: serial correlation, seasonality and nonnormality. One technological limitation, censoring, is addressed. Finally, several data record attributes that result from operational difficulties are reviewed briefly. These include missing values, different sampling frequencies, multiple observations and outliers.

B. Serial correlation

Serial correlation is generally thought of as redundancy of information between adjacent observations in a time series. It can also occur between every other observation, every third observation and so on. Serial correlation, trend and seasonality are all factors that cause data to violate the assumption of independence that underlies most statistical methods.

Characterization studies conducted on groundwater quality data from the IBM East Fishkill site indicate that serial correlation is minimal for quarterly sampling frequencies.

Specific information on how serial correlation will be handled when statistically analyzing groundwater quality data is presented in section IV.E.

C. Seasonality

Seasonality is the change in distribution of water quality variables that can be attributed to the time of year. A "season" can be any specified period of time but is generally one month (twelve seasons per year) or three months (four seasons per year). Seasonality may or may not occur as a consistent pattern.

Characterization studies conducted on groundwater quality data from the IBM East Fishkill site indicate that moderate seasonality is present in data collected from the shallow aquifer. In general, IBM has dealt with seasonality according to recommendations provided in Ward *et al.* (1988). Specific information on how seasonality will be handled when statistically analyzing groundwater quality data is presented in section IV.E.

D. Nonnormality

Many groundwater quality variables display nonnormally. Right-skewed distributions are quite common because water quality data has a lower bound of zero with many values near the detection limit.

Characterization studies conducted on groundwater quality data from the IBM East Fishkill site indicate that the data has slight to moderate nonnormality.

Specific information on how nonnormality will be handled when statistically analyzing groundwater quality data is presented in section IV.E.

E. Censoring

IBM has not censored laboratory data since 1986. This noncensoring approach is recommended by Porter (1986), Porter *et al.* (1988), Ward *et al.* (1988) and ASTM Method D4210-83. The laboratory has been asked to report the actual readings along with an estimate of measurement error. The uncensored laboratory measurements are used for all data calculations and the results are rounded to the appropriate number of significant figures prior to reporting.

F. Data Record Attributes Resulting From Operational Difficulties

Data record attributes resulting from operational difficulties have decreased considerably since groundwater monitoring began at the East Fishkill facility. Those attributes that do occur are handled as described here.

A great deal of effort is made at the IBM East Fishkill facility to avoid missing values because of the interference they cause with data analysis methods. It is IBM's policy to not replace missing values with any type of numerical response. If missing values appear to have a significant impact on a particular data analysis method, an appropriate note will accompany the results.

Although changing sampling frequencies occasionally occur at the IBM site, they are not expected to affect any of the data analysis procedures recommended in this DAP.

Multiple observations obtained for quality control purposes are stored and analyzed with the rest of the data. If a single response is needed for that time period, the multiple observations are averaged. Quality control data is not discarded unless there is clear evidence that it is invalid or there is a QC problem. Similarly, outliers are not discarded unless evidence can be found to invalidate the result.

IV. DATA ANALYSIS METHODS

A. General

Several questions dealing with information goals, data characteristics and data analysis method characteristics were addressed during the process of choosing data analysis methods. Some of the questions are listed here:

Information Goals

- What do we want to know?
- How soon do we need to know it?
- What degree of accuracy is appropriate?

Data Characteristics

- What are the attributes of the data?
- What is the sample size?
- How much historical data do we have?

Characteristics of the Data Analysis Method

- Is the data analysis method easy to understand?
- Is it widely accepted?
- Is it correct?

Data analysis methods that were chosen are discussed below. Advantages and limitations of the techniques are summarized and the rational used to choose them is described. Technical details, such as calculation procedures, are presented in Appendix E. For purposes of the following discussion, the methods are divided into four categories: (1) computational, (2) graphical, (3) direct comparison with fixed limits and (4) statistical.

B. Computational Methods

Strictly computational procedures (i.e., those involving calculations but not statistics) were chosen for information goal #4: *to determine the mass of contaminants removed at each remediation area*.

C. Graphical Methods

Plots of concentration *versus* time (i.e., time series plots) were selected to analyze data for information goal #5: *to characterize contaminant concentrations over time within the plumes*. This qualitative, graphical approach was chosen because it provides information on general patterns of contaminant behavior. Estimation of trend magnitude or determination of the statistical significance of trends is not needed at this time. Quantitative trend analysis may be appropriate in the future when data records are longer, and when the relationship between remediation effectiveness and trend is better understood. Current data collection practices (fixed interval sampling, no censoring and virtually no missing values) produce the type of data that is needed for quantitative trend analysis.

Data collected for information goal #6: *to determine the hydraulic effectiveness of remediation*, is to be converted into information by plotting the water levels on a site map and constructing contours of the water table. Another method for analyzing water elevation measurements is currently being developed. This method uses vector analysis.

D. Direct Comparison With Fixed Limits

Direct comparison with fixed limits was selected as a technique to analyze data for five of the twelve information goals. The method involves comparison of individual concentration values to fixed limits without the use of statistics.

The five information goals for which direct comparison will be used are listed in Table 1. Fixed limits that have been set for each goal are also listed.

The method of direct comparison with fixed limits was chosen for the information goals listed in Table 1 because it was felt that action should be taken if a single verified value falls above the specified limits. Recommended action for each information goal is described in Appendix A.

TABLE 1. Information goals and fixed limits for direct comparison.

INFORMATION GOAL		FIXED LIMITS
#1:	To ensure adequacy of the treatment facility by screening for the presence of new compounds.	MDL
#7:	To measure water quality down-gradient of AOCs where flow may pass across a site boundary*.	2 x MDL
#8:	To confirm previous monitoring results that indicate that contaminants are not migrating offsite from onsite.	GWPS
#9:	To confirm previous monitoring results that indicate that contaminants are not migrating onsite from offsite.	GWPS
#10:	To ensure that all chemicals of concern are properly identified and monitored.	MDL

* Interval estimation is the primary procedure chosen for information goal #7 (see the "statistical methods" section). Comparison with fixed limits was chosen as a secondary method to ensure that very high values are dealt with immediately.

E. Statistical Methods

E.1 Control Charts

Control chart methods were chosen to analyze data for information goal #2: *to ensure adequacy of the treatment facility by detecting changes in input concentrations*.

Control charts are a graphical representation of a statistical quality control procedure. They consist of a horizontal line corresponding to the average value of the characteristic in question, together with upper and/or lower control limits (Marriott, 1990). Most commonly, the horizontal axis is in units of time and the vertical axis isin units of standard deviation. Sampling results or statistics of sampling results are plotted sequentially on the chart. If a plotted point falls outside the limits, the process is said to be out of control.

"Out of control" simply means that statistically unusual variation from normal performance has been detected (Berthouex *et al.*, 1978). Action taken in response to an out of control signal may consist of adjusting manufacturing equipment, increasing the retention time in a clarifier, or just taking a closer look at the data.

Control charts have traditionally been used for process quality control in the manufacturing industry. In the field of water quality management, control charts are used to monitor sewage treatment plant effluent (Berthouex and Hunter, 1975; Berthouex *et al.*, 1978 and Vaughan and

Russell, 1983) and to analyze water quality data at RCRA sites (U.S. EPA, 1989 and Starks and Flatman, 1991).

Environmental quality data often violates the underlying assumptions of control charts by exhibiting serial correlation, seasonality, nonconstant variance and nonnormal distributions (Berthouex *et al.*, 1978). Complications arising from "messy" data are exacerbated by the small sample sizes encountered in groundwater quality monitoring. Research that addresses these issues is still in its early stages. The results of control chart analysis of groundwater quality data are therefore subject to a high degree of uncertainty and should be used accordingly. It would be inappropriate to base major environmental decisions on the results of control chart analysis.

If their limitations are acknowledged, however, control charts can be a effective tool for quickly detecting shifts in water quality. They are versatile and easy to use. Most importantly, they provide an ongoing visual account of water quality. Research has shown that shifts in quality are often more readily apparent on CUSUM control charts than on time plots (Hockman and Lucas, 1987).

Results of control chart analyses will be used as a tool to optimize the efficiency of IBM groundwater treatment facilities. Results will not be used as the basis for regulatory action.

The decision was made to employ a control chart procedure known as the "combined Shewhart-CUSUM method" (Lucas, 1982) to analyze data for information goal #2. Large shifts in concentration values are detected by Shewart limits, whereas small changes that persist are detected by CUSUM limits. The combined Shewhart-CUSUM method is easier to apply than other control chart methods that provide similar results (Lucas, 1982). Details of the procedure are presented in Appendix E-1.

Evaluation of existing data indicates that low to moderate amounts of serial correlation, seasonality and nonnormality should be expected in the data which is to be analyzed by control chart methods.

Serial correlation and seasonality increase the probability of Type I error (i.e., false positives) (Loftis *et al.*, 1987 and U.S. EPA, 1989). The effects of nonnormality depend on the actual distribution that is present (Schilling and Nelson, 1976).

No special adjustments were made to the combined Shewhart-CUSUM method to account for serial correlation, seasonality or nonnormality. This decision was based on three factors: (1) adjustments may complicate the method making it more difficult to understand from an intuitive viewpoint, (2) modifications that involve data transformations can decrease information content of the data and/or make visual inspection of the data less effective and

(3) modifications to account for data record attributes may affect the results of control charts in ways that are not well understood.

E.2 Interval Estimation of the Difference in Medians

Interval estimation was one of the methods chosen *to determine the lateral extent of contamination plumes* (information goal #3). An interval estimate of the difference between medians for the current two years of data and the previous two years of monitoring data will be made for each key compound (or for each sum of key compounds). The purpose of the estimate is to provide the person who is drawing plumes with a feel for if concentrations are changing enough in the well to warrant shifting the position of the plume boundary. (See the DAP summary sheet for information goal #3 in Appendix A for a more detailed explanation of how plume boundaries will be drawn.)

An interval estimate is an intuitively appealing way of presenting information. The width of the confidence interval indicates how much reliance should be placed on the estimate. Similar information can be obtained from the yes/no results of a hypothesis test accompanied by the appropriate operating characteristic curve (Natrella, 1972). Operating characteristics curves, however, are difficult to construct and interpret.

The particular estimator chosen for information goal #3 is the Hodges-Lehmann estimator (Hodges and Lehmann, 1963). It is simply the median of all pairwise differences between the two groups. Details of the procedure are given in Appendix E-2.

The Hodges-Lehmann estimator was chosen because it is robust in the presence of serial correlation, seasonality and serial correlation, all of which may be present in small to moderate amounts (Hirsch, 1988).

E.3 Interval Estimation of the Median

Interval estimation of the median was the method selected to analyze data for information goal #7: *to measure water quality down-gradient of AOCs where flow may pass across a site boundary*. Interval estimates of the median will be made on a yearly basis and compared to the groundwater protection standard (GWPS). Action will be taken if the entire interval is above the GWPS.

A statistical approach was chosen instead of direct comparison because some of the compounds are expected to be present at levels above the MDL. Direct comparison is appropriate when the main objective is to detect compounds that have not been found previously in a particular well. The objective here, however, is to determine if any compounds are present at levels above the GWPS, while accounting for distributional variability.

Details of the interval estimation procedure are presented in Appendix E-3.

V. CRITERIA FOR CHOOSING COMPOUNDS FOR LABORATORY ANALYSIS, DATA SUMMARIZATION AND DATA ANALYSIS

A. Routinely Monitored Compounds

All samples are analyzed for the "routinely monitored compounds" which are listed in Table C-2. Results are reported as raw data.

B. Selected Compounds

A procedure that employed boxplots was used to decide which of the routinely monitored compounds should be designated as "selected compounds." Boxplots were constructed for data from the second quarter of '89 through the first quarter of '91. The data represented either single compounds or groups of compounds. Compounds (or groups of compounds) whose upper fence fell above the groundwater protection standard (usually 5μ/L) were designated as selected compounds. Species were also designated as selected compounds if any single verified value exceeded the groundwater protection standard.

Selected compounds are listed in Table C-3. Results of analyses of selected compounds are reported as summary statistics.

D. Key Compounds

Key compounds are those routinely monitored compounds which are chosen for special data analysis techniques. Key compounds vary according to area as shown in Table C-4. They were chosen based on concentration, importance and their ability to meet information goals.

VI. CRITERIA FOR CHOOSING SAMPLING INTERVALS

Factors that were considered when choosing sampling intervals for this data analysis protocol include:

- Regulatory requirements and guidance, which generally recommend a minimum of semi-annual sampling.
- Monitoring and statistical information goals.
- Historical water quality record for well and adjacent points.
- Location of well and direction of groundwater flow.
- Risk to human health and the environment.

VII. PROCEDURES FOR REVIEWING THE DATA ANALYSIS PROTOCOL

The DAP will be reviewed every two years. It will be reviewed more frequently if: (1) new regulations are promulgated that affect the content of the DAP, (2) it becomes evident that a section of the DAP is invalid or (3) a particular revision would greatly improve the effectiveness of the DAP. Minor revisions will be saved for the next scheduled review.

Protocol review will be overseen by the groundwater quality program manager. The manager will sign-off on all revisions.

Factors that will be considered during the reviews include current research, amended regulations and new information obtained from recent monitoring data.

All revisions made to the data analysis protocol will be thoroughly documented and, if necessary, reported to the NYSDEC. Each revision will have the following heading:

GROUNDWATER QUALITY DATA ANALYSIS PROTOCOL REVISION

Date:

Author:

Approved by:

Summary of Revision:

All copies of the DAP will be kept in loose-leaf notebooks to facilitate the revision process. Revision sheets will be placed at the back of the protocol. Sections of the protocol that are no longer applicable will be crossed out but not removed.

With a loose-leaf format, the potential exists for sheets to be lost or disorganized. To remedy this situation, one copy of the DAP will be designated as the master copy. It will be carefully maintained and kept in one location to serve as a reference to individuals who have questions regarding organization of their own copies of the DAP.

VIII. REFERENCES

Bell, H.F. and H.P. DeLong. 1988. Data characteristics: groundwater monitoring's "Catch 22." American Chemical Society--75 Years of Progress. Preprints of papers presented at the 196th ACS national meeting, Los Angeles, CA, September 25039, 1988. 28(2):20-24,

Berthouex, P.M., W.G. Hunter and L. Pallesen. 1987. Monitoring sewage treatment plants: some quality control aspects. Journal of Quality Technology. 10(4):139-149.

Berthouex, P.M. and W.G. Hunter. 1975. Treatment plant monitoring programs: a preliminary analysis. Journal Water Pollution Control Federation. 47(8):2143-2156.

Bissell, A.F. 1984. The performance of control charts and Cusums under linear trend. Applied Statistics. 33(2):145-151.

Cheremisinoff, N.P. 1988. Computer statistics for environmental engineers. Pollution Engineering. 20(1):54-62.

Hirsch, R.M. 1988. Statistical methods and sampling design for estimating step trends in surface-water quality. Water Resources Bulletin. 24(3):493-503.

Hockman, K.K. and J.M. Lucas. 1987. Variability reduction through subvessel CUSUM control. Journal of Quality Technology. 19(3):113-121.

Hodges, J.L. and E.L. Lehmann. 1963. Estimates of location based on rank tests. The Annals of Mathematical Statistics. 34:598-611.

Loftis, J.C., P.S. Porter and G. Settembre. 1987. Statistical analysis of industrial wastewater monitoring data. Journal Water Pollution Control Federation. 59(3):145-151.

Lucas, J.M. 1982. Combined Shewhart-CUSUM quality control schemes. Journal of Quality Technology. 14(2):51-59.

Lucas, J.M. 1985. Cumulative Sum (CUSUM) control schemes. Communication in Statistics--Theory and Methods. 14(11):2689-2704.

Lucas, J.M. and R.B. Crosier. 1982. Robust CUSUM: a robustness study for CUSUM quality control schemes. Communications in Statistics--Theory and Methods. 11(23):2669-2687.

Marriott, F.H.C. 1990. A Dictionary of Statistical Terms. Longman Scientific and Technical, Essex, England, and John Wiley and Sons, Inc., New York, NY.

Natrella, M.G. 1972. The relationship between confidence intervals and tests of significance. Statistical Issues: A Reader for the Behavioral Sciences. R.E. Kirk ed., Brooks/Cole Publishing Co., Monterey, California. p.113-117.

NRC. 1990. Managing troubled waters--the role of marine environmental monitoring. A report by the Committee on a Systems Assessment of Marine Environmental Monitoring, Marine Board, Commission on Engineering and Technical Systems, National Research Council. National Academy Press, Washington D.C.

Page, E.S. 1954. Continuous inspection schemes. Biometrika. 41(1,2):100-114.

Page, E.S. 1961. Cumulative sum charts. Technometrics. 3(1):1-9.

Porter, P.S. 1986. Statistical analysis of water quality data affected by the limits of detection. Unpublished Ph.D. Thesis, Department of Agricultural and Chemical Engineering, Colorado State University, Fort Collins, Colorado.

Schilling, E.G. and P.R. Nelson. 1976. The effect of non-normality on the control limits of X charts. Journal of Quality Technology. 8(4):183-188.

Starks, T.H. 1989. Evaluation of control chart methodologies for RCRA (Resource Conservation and Recovery Act) Waste Sites. Office of Research and Development, U.S. Environmental Protection Agency, Las Vegas, Nevada. U.S. EPA Report No. EPA/600/4-88/040.

Starks, T.H. and G.T. Flatman. 1991. RCRA ground-water monitoring decision procedures viewed as quality control schemes. Environmental Monitoring and Assessment. 16:19-37.

U.S. EPA. 1989. Statistical analysis of ground-water monitoring data at RCRA facilities--interim final guidance. Office of Solid Waste, Waste Management Division, U.S. Environmental Protection Agency, Washington, D.C., February.

Vasilopoulos, A.V. and A.P. Stamboulis. 1978. Modification of control chart limits in the presence of data correlation. Journal of Quality Technology. 10(1):20-29.

Vaughan, W.J. and C.S. Russell. 1983. Monitoring point sources of pollution: answers and more questions from statistical quality control. The American Statistician. 37(4):476-487.

Ward, R.C., J.C. Loftis and G.B. McBride. 1990. Design of Water Quality Monitoring Systems. Van Nostrand Reinhold Company, New York.

Wurman, R.S. 1989. Information Anxiety. Doubleday, New York, New York.

APPENDIX A--DATA ANALYSIS PROTOCOL SUMMARY SHEETS

INFORMATION GOAL #1: To ensure adequacy of the treatment facility by screening for the presence of new[1] compounds.

COMPOUNDS TO BE ANALYZED: "Appendix 33" compounds (Table C-1).

WELL TYPE: Extraction wells

NUMBER OF WELLS TO BE SAMPLED: 7

SAMPLING FREQUENCY: 1/year for three years

DATA ANALYSIS PROCEDURES: Concentrations will be compared with the method detection limit (MDL) as described below.

ACTIONS TO BE TAKEN BASED ON LABORATORY RESULTS:

If a new compound is detected at or above the MDL, the well will be resampled. If the concentration is still at or above the MDL, then the presence of the compound is confirmed[2] and:

(1) The pertinent well, compound and concentration will be reported to the New York State Department of Environmental Conservation (NYSDEC) within seven days of confirmation.

(2) If necessary, adjustments will be made to the treatment system to account for the presence of the new compound.

(3) If the new compound is not included in the list of routinely monitored compounds (Table C-2), a decision will be made as to whether or not it should be added to the list. The decision will be based on several factors including the concentration of the compound in relation to the GWPS, and the location and direction of groundwater flow.

REPORTING:

The following information will be included in the first semi-annual report that is written subsequent to sampling:

---

[1] "New" compounds are those that have not previously been detected and confirmed in a particular well.

[2] Organics will be analyzed by GC/MS where applicable in order to yield positive qualitative identification.

(1) Results of all laboratory analyses will be reported in a raw data format.

(2) If a resample is conducted and the compound of interest is no longer at or above the MDL, then the original analysis will be labeled "invalid."

(3) If a resample is conducted and the presence of the compound is confirmed, both analyses will be considered valid and the pertinent well, compound and concentrations will be reported separately from the raw data.

INFORMATION GOAL #2: To ensure adequacy of the treatment facility by detecting changes in input concentrations.

COMPOUNDS TO BE ANALYZED: Routinely monitored compounds (Table C-2).

WELL TYPE: Extraction wells

NUMBER OF WELLS TO BE SAMPLED: 12

SAMPLING FREQUENCY: 12/year

DATA ANALYSIS PROCEDURES:

All key compounds (Table C-4) will be tracked by control chart methods[3] subsequent to each sampling episode (i.e., 12 times per year). If concentrations are found to be "out of control", the data will be inspected more closely and, if necessary, adjustments will be made to the treatment system to account for increased concentrations of the compound.

ACTIONS TO BE TAKEN BASED ON LABORATORY RESULTS:

Concentrations of all key compounds will be tracked by control chart methods as described above.

REPORTING:

The following information will be included in the first semi-annual report that is written subsequent to sampling:

(1) Results of all laboratory analyses will be reported as raw data.

(2) Summary statistics from analyses of selected compounds (Table C-3) will be presented.

(3) Control charts constructed from concentrations of key compounds will be presented as shown in Figure D-1.

(4) Any adjustments made to the treatment system resulting from the presence of "out of control" compounds will be discussed.

(5) If any compounds are added to the list of key compounds, the data and rational supporting the addition will be supplied.

---

[3] Details of control chart methods that will be used to track concentrations for process control are presented in Appendix E-I.

INFORMATION GOAL #3: To determine the lateral extent of contamination plumes.

COMPOUNDS TO BE ANALYZED: Routinely monitored compounds (Table C-2).

WELL TYPE: Wells near plume perimeters.

NUMBER OF WELLS TO BE SAMPLED: 87

SAMPLING FREQUENCY: 2,4, or 12 per year.

DATA ANALYSIS PROCEDURES:

The following procedures will be conducted on a yearly basis:

(1) Annual medians of key compounds (Table C-4) (or sums of key compounds) will be plotted on a site map.

(2) An interval estimate of the difference between medians for the current two years of data and the previous two years of monitoring data will be made for each key compound (or for each sum of key compounds). The purpose in doing this is to give the person who is drawing the plumes a feel for if concentrations are changing enough in the well to warrant shifting the position of the plume boundary. Details of this statistical procedure are presented in Appendix E-II.

(3) Plume boundaries will be drawn based on the above information and on personal judgement. Knowledge of flow direction, site hydrogeology and contaminant transport behavior should support "personal judgement" decisions.

ACTIONS TO BE TAKEN BASED ON LABORATORY RESULTS: None

REPORTING:

The following information will be included in each annual report:

(1) Results of all laboratory analyses will be reported as raw data.

(2) Summary statistics from analyses of selected compounds (Table C-3) will be presented.

(3) Annual medians of key compounds (or sums of key compounds) will be plotted on a site map.

(4) Interval estimates of the differences in medians between the current two years data and the previous two years data will be given for each key compound (or for each sum of key compounds) (see Figure D-2).

(5) Plume boundaries for all six Areas of Concern (AOCs) will be drawn on a site map. Plume boundaries are defined as an annual median concentration equal to 5µ/L of one key compound or sum of key compounds.

INFORMATION GOAL #4: To determine the mass of contaminants removed at each remediation area.

COMPOUNDS TO BE ANALYZED: Routinely monitored compounds (Table C-2).

WELL TYPE TO BE SAMPLED: Extraction wells

NUMBER OF WELLS TO BE SAMPLED: 11

SAMPLING FREQUENCY: 12/year

DATA ANALYSIS PROCEDURES: The volume of water pumped will be multiplied by the concentrations of each key compound (Table C-4) to obtain the mass of contaminants removed.

ACTIONS TO BE TAKEN BASED ON LABORATORY RESULTS: None

REPORTING:

The following information will be included in each annual report:

(1) Results of all laboratory analyses will be reported as raw data.

(2) Summary statistics from analyses of selected compounds (Table C-3) will be presented.

(3) Masses of contaminants removed, as well as volume and concentration data used calculate the masses, will be reported.

INFORMATION GOAL #5: To characterize contaminant concentrations over time within the plumes.

COMPOUNDS TO BE ANALYZED: Routinely monitored compounds (Table C-2).

WELL TYPE TO BE SAMPLED: Wells within the plumes.

NUMBER OF WELLS TO BE SAMPLED: 11

SAMPLING FREQUENCY: 2 or 12 per year.

DATA ANALYSIS PROCEDURES: Plots of time *versus* concentration will be constructed for all key compounds (Table C-4). The plots will be observed in order to improve current understanding of the relationship between remediation efforts and contaminant behavior.

ACTIONS TO BE TAKEN BASED ON LABORATORY RESULTS: None

REPORTING:

The following information will be included in each annual report:

(1) Results of all laboratory analyses will be reported as raw data.

(2) Summary statistics from analyses of selected compounds (Table C-3) will be presented.

(3) Plots of time *versus* concentration for key compounds will be provided (see Figure D-3).

INFORMATION GOAL #6: To determine the hydraulic effectiveness of remediation.

COMPOUNDS TO BE ANALYZED: Not applicable because only water elevations will be measured.

WELL TYPE TO BE SAMPLED: All onsite (i.e., Main Site and West Complex) monitoring wells.

NUMBER OF WELLS TO BE SAMPLED: 149

SAMPLING FREQUENCY: 4/year

DATA ANALYSIS PROCEDURES: Water elevations will be used to construct contour maps of the water table. Pumping may be adjusted based on the contour maps.

ACTIONS TO BE TAKEN BASED ON LABORATORY RESULTS: None

REPORTING:

The following information will be included in each annual report.

(1) All water elevations measured in monitoring wells will be reported in a raw data format.

(2) Contour plots of water table elevations will be provided.

INFORMATION GOAL #7: To measure water quality down-gradient of AOCs where flow may pass across a site boundary.

COMPOUNDS TO BE ANALYZED: Routinely monitored compounds (Table C-2).

WELL TYPE TO BE SAMPLED: Wells that are both near the property boundary and down-gradient of an AOC.

NUMBER OF WELLS TO BE SAMPLED: 6

SAMPLING FREQUENCY: 4 or 12 per year.

DATA ANALYSIS PROCEDURES:

On an annual basis, concentrations will be compared to the Ground Water Protection Standard (GWPS) using interval estimation. Details of the procedure are presented in Appendix E-III.

If data analysis procedures indicate that the GWPS has been exceeded, IBM will attempt to determine the reason for the change in groundwater quality as well as the implications of the change in terms of protection of human health and the environment. Corrective action will be taken if necessary.

ACTIONS TO BE TAKEN BASED ON LABORATORY RESULTS:

If a compound is detected that is at least twice the level of the GWPS, the well will be resampled. If the concentration is still at or above twice the GWPS, then the concentration is confirmed and:

(1) The pertinent well, compound and concentration will be reported to the NYSDEC within seven days of confirmation.

(2) IBM will attempt to determine the reason for the change in groundwater quality, and the implications of the change in terms of protection of human health and the environment. Corrective action will be taken if necessary.

REPORTING:

The following information will be included in each annual report:

(1) Results of all laboratory analyses will be reported as raw data.

(2) Summary statistics from analyses of selected compounds (Table C-3) will be presented.

(3) If a resample is conducted and the compound of interest is no longer at or above twice the GWPS, then the original analysis will be labeled "invalid."

(4) If a resample is conducted and the concentration is confirmed, both analyses will be considered valid and the pertinent well, compound and concentrations will be reported separately from the raw data.

(5) Interval estimates of annual concentration medians will be given for all routinely monitored compounds (see Figure D-4).

(6) Documentation will be provided of any actions, decisions or observations made as the result of either: (a) the discovery of interval estimates of annual medians that exceed the GWPS or (b) the discovery of concentrations above twice the GWPS.

INFORMATION GOAL #8: To confirm previous monitoring results that indicate that contaminants are not migrating offsite from onsite.

COMPOUNDS TO BE ANALYZED: Routinely monitored compounds (Table C-2).

WELL TYPE TO BE SAMPLED: Clean wells located near the property boundary.

NUMBER OF WELLS TO BE SAMPLED: 14

SAMPLING FREQUENCY: 2/year

DATA ANALYSIS PROCEDURES: Concentrations will be compared with the GWPS as described below.

ACTIONS TO BE TAKEN BASED ON LABORATORY RESULTS:

If a compound is detected at or above the GWPS, the well will be resampled. If the concentration is still at or above the GWPS, then the concentration is confirmed and:

(1) The pertinent well, compound and concentration will be reported to the NYSDEC within seven days of confirmation.

(2) IBM will immediately attempt to determine the reason for the change in groundwater quality, and the implications of the change in terms of protection of human health and the environment. Corrective action will be taken if necessary.

(3) If appropriate, the well may be redesignated as a well for meeting goal #7.

REPORTING:

The following information will be included in each semi-annual report:

(1) Results of all laboratory analyses will be reported as raw data.

(2) If a resample is conducted and the compound of interest is no longer at or above the GWPS, then the original analysis will be labeled "invalid."

(3) If a resample is conducted and the presence of the compound is confirmed, both analyses will be considered valid and the pertinent well, compound and

concentrations will be reported separately from the raw data.

(4) Any actions, decisions or observations made as the result of discovery of concentrations above the GWPS will be reported.

(5) Summary statistics from analyses of selected compounds (Table C-3) will be presented.

INFORMATION GOAL #9: To confirm previous monitoring results that indicate that contaminants are not migrating onsite from offsite.

COMPOUNDS TO BE ANALYZED: Routinely monitored compounds (Table C-2).

WELL TYPE TO BE SAMPLED: Clean wells located near the property boundary.

NUMBER OF WELLS TO BE SAMPLED: 8

SAMPLING FREQUENCY: 2/year

DATA ANALYSIS PROCEDURES: Concentrations will be compared with the GWPS as described below.

ACTIONS TO BE TAKEN BASED ON LABORATORY RESULTS:

If a compound is detected at or above the GWPS, the well will be resampled. If the concentration is still at or above the GWPS, then the concentration is confirmed and:

(1) The pertinent well, compound and concentration will be reported to the NYSDEC within seven days of confirmation.

(2) IBM will immediately attempt to determine the reason for the change in groundwater quality, and the implications of the change in terms of protection of human health and the environment.

REPORTING:

The following information will be included in each semi-annual report:

(1) Results of all laboratory analyses will be reported as raw data.

(2) If a resample is conducted and the compound of interest is no longer at or above the GWPS, then the original analysis will be labeled "invalid."

(3) If a resample is conducted and the presence of the compound is confirmed, both analyses will be considered valid and the pertinent well, compound and concentrations will be reported separately from the raw data.

(4) Any actions, decisions or observations made as the result of discovery of concentrations above the GWPS will be reported.

(5) Summary statistics from analyses of selected compounds (Table C-3) will be presented.

INFORMATION GOAL #10: To ensure that all chemicals of concern are properly identified and monitored.

COMPOUNDS TO BE ANALYZED: "Appendix 33" compounds (Table C-1).

WELL TYPE TO BE SAMPLED: Wells within the plumes.

NUMBER OF WELLS TO BE SAMPLED: 8

SAMPLING FREQUENCY: 1/year for 3 years.

DATA ANALYSIS PROCEDURES: Concentrations will be compared with the MDL as described below.

ACTIONS TO BE TAKEN BASED ON LABORATORY RESULTS:

If an additional[4] compound is detected at or above the MDL, the well will be resampled. If the concentration is still at or above the MDL, then the presence of the compound is confirmed and:

(1) The pertinent well, compound and concentration will be reported to the NYSDEC within seven days of confirmation.

(2) The well will sampled for two quarters following the original confirmation. Laboratory results will be reported to the NYSDEC.

(3) A decision will be made as to whether or not the additional compound should be placed on the list of routinely monitored compounds (i.e., Table C-2). The decision will be based on several factors including the concentration of the compound in relation to the GWPS, and the location and direction of groundwater flow.

REPORTING:

The following information will be included in the first semi-annual report that is written subsequent to sampling:

(1) Results of all laboratory analyses will be reported in a raw data format.

---

[4] "Additional" refers to compounds other than those that are routinely monitored (i.e. other than those compounds listed in Table C-2).

(2) If a resample is conducted and the compound of interest is no longer at or above the MDL, then the original analysis will be labeled "invalid."

(3) If a resample is conducted and the presence of the compound is confirmed, both analyses will be considered valid and the pertinent well, compound and concentrations will be reported separately from the raw data.

INFORMATION GOAL #11: To provide a qualitative basis for comparison with data from "Appendix 33" monitoring of contaminated wells.

COMPOUNDS TO BE ANALYZED: "Appendix 33" compounds (Table C-1).

WELL TYPE TO BE SAMPLED: Clean wells located away from AOCs.

NUMBER OF WELLS TO BE SAMPLED: 2

SAMPLING FREQUENCY: 1/year for 3 years.

DATA ANALYSIS PROCEDURES: To be determined.

ACTIONS TO BE TAKEN BASED ON LABORATORY RESULTS: None

REPORTING:

The following information will be included in the first semi-annual report that is written subsequent to sampling:

(1) Results of all laboratory analyses will be reported in a raw data format.

INFORMATION GOAL #12: To provide data for investigative studies associated with the NYS Part 373 Corrective Action Program.

COMPOUNDS TO BE ANALYZED: Routinely monitored compounds (Table C-2).

WELL TYPE TO BE SAMPLED: Wells in areas under investigation.

NUMBER OF WELLS TO BE SAMPLED: 64

SAMPLING FREQUENCY: 4/year

DATA ANALYSIS PROCEDURES: To be determined[5]

ACTIONS TO BE TAKEN BASED ON LABORATORY RESULTS: None

REPORTING:

The following information will be included in the first semi-annual report that is written subsequent to sampling:

(1) Results of all laboratory analyses will be reported as raw data.

(2) Summary statistics from analyses of selected compounds (Table C-3) will be presented.

(3) A summary of project status will be provided.

(4) Annual medians of key compound(s) will be plotted on a site map.

---

[5] A separate data analysis protocol will be written for each Remedial Facilities Investigation (RFI) work plan.

APPENDIX B--SAMPLING FREQUENCIES

TABLE B-1. Sampling frequencies (per year)--area A remediation area.

Well	Information Goal #											
	1	2	3	4	5	6	7	8	9	10	11	12
702			2			4						
007			2			4						
998			2			4						
999			2			4						
720			2			4						
761			2			4						
99			2			4						
763			2			4						
204			2			4						
714						4				1		
Sump	1	12				4						
16				2	2	4						
103				2	2	4						
104				2	2	4						
744						4						4
745						4						4
15						4						4

TABLE B-1. Sampling frequencies (per year)--area A remediation area, cont.

Well	Information Goal #											
	1	2	3	4	5	6	7	8	9	10	11	12
706						4				1		
SW						4						12

TABLE B-2. Sampling frequencies (per year)--area B remediation area.

Well	Information Goal #											
	1	2	3	4	5	6	7	8	9	10	11	12
863	1	12	12	12	12	4						
834			4			4						
571			4			4						
064			4			4	4					
065			4			4	4					
863 GAC		12				4						

TABLE B-3. Sampling frequencies (per year)--area C remediation area.

Well	Information Goal #											
	1	2	3	4	5	6	7	8	9	10	11	12
563			2			4						
561			2			4						
559			2			4				1		
053			2			4						
054			2			4						
932			2			4						
944			2			4						
943			2			4						
069			2			4						

TABLE B-4. Sampling frequencies (per year)--area D remediation area.

Well	Information Goal #											
	1	2	3	4	5	6	7	8	9	10	11	12
032	1	12	12	12	12	4						
757			4			4						
067			4			4	4					
066			4			4	4					

TABLE B-5. Sampling frequencies (per year)--SEQ remediation area.

Well	Information Goal #											
	1	2	3	4	5	6	7	8	9	10	11	12
048			4			4						
552			4			4				1		
543			4			4						
545			4			4						
990			12			4	12					
987			12			4	12					
520			4			4						
502			4			4						
503			4			4						
859						4						4
860						4						4
861						4						4
862						4						4

TABLE B-6. Sampling frequencies (per year)--area C B/330 investigative area.

Well	Information Goal #											
	1	2	3	4	5	6	7	8	9	10	11	12
574						4						1
178						4						4
576						4						4
921						4						2
922						4						2
923						4						2
924						4						2
954						4						2
955						4						2
956						4						2
957						4						2
965						4						2
966						4						2
941						4						2
PW-7	1	12	12	12	12	4						
572						4						1
181						4				1		4

TABLE B-6. Sampling frequencies (per year)--area C B/330 investigative area, cont.

Well	Information Goal #											
	1	2	3	4	5	6	7	8	9	10	11	12
756						4						4
062						4						4
063						4						4
061						4						4
059						4						4
060						4						4
143						4						2
145						4						2
146						4						2
929						4						2
930						4						2
931						4						2
953						4						2

TABLE B-7. Sampling frequencies (per year)--building 322 investigative area.

Well	Information Goal #											
	1	2	3	4	5	6	7	8	9	10	11	12
735						4						4
740						4						4
766						4						4
767						4						4
768						4						4
769						4						4
770						4						4
771						4						4
772						4						4
773						4						4
754						4						4
755						4						4
792						4						4
793						4						4
781						4						4
203						4						4
151						4						4

TABLE B-7. Sampling frequencies (per year)--building 322 investigative area, cont.

Well	Information Goal #											
	1	2	3	4	5	6	7	8	9	10	11	12
949						4						4
743						4						4
161						4						4
35						4						4
68						4						4
774						4						4
775						4						4
791						4						4
789						4						4
202						4						4

TABLE B-8. Sampling frequencies (per year)--bedrock aquifer.

Well	Information Goal #											
	1	2	3	4	5	6	7	8	9	10	11	12
PW-1		12	12	12	12	4						
PW-2	1	12	12	12	12	4						
PW-4	1	12	12	12	12	4						
PW-5/5A		12	12	12	12	4						
PW-6		12	12			4						
PW-7	1	12	12	12	12	4						
PW-9 (when on)		12	12	12	12	4						
GAC-Input	1	12				4						
113			4			4						
535			4			4						
958			4			4						
515			4			4						
963			4			4						
964			4			4						
521			4			4						

TABLE B-8. Sampling frequencies (per year)--bedrock aquifer, cont.

Well	Information Goal #											
	1	2	3	4	5	6	7	8	9	10	11	12
857			4			4						
858			4			4						
166			2			4			2			
167			2			4			2			
087			2			4			2			
085			2			4			2			
722			2			4						
711			2			4						
105			2			4						
106			2			4						
704			2			4						
108			2			4						
109			2			4						
959			2			4				1		
965			2			4						
966			2			4						
564			2			4						

TABLE B-8. Sampling frequencies (per year)--bedrock aquifer, cont.

Well	Information Goal #											
	1	2	3	4	5	6	7	8	9	10	11	12
563			2			4						
559			2			4						
932			2			4						
944			2			4						
943			2			4						
145			2			4						
146			2			4						
929			2			4						
930			2			4						
931			2			4						
941			2			4				1		
953			2			4						
143			2			4						
156			2			4						
945			2			4						
952			2			4						
737			2			4						

TABLE B-8. Sampling frequencies (per year)--bedrock aquifer, cont.

Well	Information Goal #											
	1	2	3	4	5	6	7	8	9	10	11	12
742			2			4						
173			2			4						
150			2			4		2				
779			2			4		2				
778			2			4		2				
777			2			4		2				
739			2			4						
716			2			4				1		

TABLE B-9. Sampling frequencies (per year)--west complex.

Well	Information Goal #											
	1	2	3	4	5	6	7	8	9	10	11	12
407											1	
408											1	

Note: Monitoring wells for information goal #8 were selected after the initial submittal of the groundwater monitoring proposal and are therefore not included in the Appendix B tables.

APPENDIX C--CATEGORIES OF COMPOUNDS

TABLE C-1. Appendix 33 compounds.

Appendix 33, found in **6 NYCRR Subpart 373-2**[1], is a list of 17 inorganic compounds and 202 organic compounds. Various sections of the Subpart 373-2 regulations require that groundwater samples be analyzed for the presence of Appendix 33 compounds. The regulations also describe actions that should be taken if Appendix 33 compounds are confirmed to be present. Appendix 33 provides the compound's chemical name, the Chemical Abstracts Service Registry Number, the Chemical Abstracts Service Index Name, suggested analytical methods and Practical Quantitation Limits (PQL's). Regulatory requirements pertain only to the list of substances. The analytical methods and PQL's are given solely for information purposes.

[1] New York State Department of Environmental Conservation, *Final Status Standards for Owners and Operators of Hazardous Waste Treatment, Storage and Disposal Facilities*, Albany, NY, January 31, 1992.

TABLE C-2. Routinely monitored compounds.

Constituents	CAS #
Acetone	67-64-1
Benzene[2]	71-43-2
Bromochloromethane	74-97-5
Chlorobenzene[2]	108-90-7
Chloroethane	75-00-3
Chloroform	67-66-3
Chloromethane	74-87-3
Dibromochloromethane	124-48-1
1,2-Dibromoethane	106-93-4
1,2-Dichlorobenzene[2]	95-50-1
1,3-Dichlorobenzene[2]	541-73-1
1,4-Dichlorobenzene[2]	106-46-7
Dichlorodifluoromethane (Freon 12)[2]	75-71-8
1,1-Dichloroethene[2]	75-35-4
cis-1,2-Dichloroethene[2]	156-59-2
Trans-1,2-Dichloroethene	156-60-5
1,3-Dichloropropane	142-28-9
Ethylbenzene[2]	100-41-4
Trichlorotrifluoroethane (Freon 113 or Freon TF)[2]	76-13-1
Hexachlorobutadiene	87-68-3
Isopropylbenzene	98-82-8
p-Isopropyltoluene	99-878-6
Methylene Chloride[2]	75-09-2
Methyl-t-Butyl Ether	1634-04-4
Naphthalene	106-44-5

[2] Listed in Table C-3.

TABLE C-2. Routinely monitored compounds, cont.

Constituents	CAS #
n-Propylbenzene	103-65-1
1,1,2,2-Tetrachloroethene[3]	127-18-4
THF	109-99-9
Toluene	108-88-3
1,2,3-Trichlorobenzene[3]	95-63-6
1,2,4-Trichlorobenzene[3]	96-18-4
1,1,2-Trichloroethene	79-01-6
Trichlorofluoromethane (Freon 11)	75-69-4
1,2,4-Trimethylbenzene	95-63-6
1,3,5-Trimethylbenzene	108-67-8
Vinyl Chloride[3]	75-01-4
m-Xylene	108-38-3
p-Xylene	106-42-3
o-xylene	95-47-6
1,1,1-Trichloroethane	71-55-6
n-Butylbenzene	104-51-8
sec-Butylbenzene	135-98-8
1,1-Dichloroethane	75-34-3
1,2-Dichloroethane	107-06-2
1,2-Dichloro-1,2,2-Trifluoroethane (Freon 123-A)[3]	354-23-4
Chromium (total)[3]	7440-47-3
Fluoride[3]	7782-41-4[4]
Zinc (total)[3]	7440-66-6

[3] Listed in Table C-3.

[4] CAS # for Fluorine.

Table C-3. Selected compounds.

Constituents	Method[5]	Conc. Limit (ug/L)[6]
Benzene	8021	0.7
Chlorobenzene	8021	5.0
1,2-Dichlorobenzene[7]	8021	4.7
1,2,3-Trichlorobenzene	8021	5.0
1,2,4-Trichlorobenzene	8021	5.0
1,3-Dichlorobenzene	8021	5.0
1,4-Dichlorobenzene[7]	8021	4.7
Dichlorodifluoromethane (Freon 12)	8021	5.0
1,1-Dichloroethene	8021	5.0
Methylene Chloride	8021	5.0
Ethylbenzene	8021	5.0
Vinyl Chloride	8021	2.0
1,2-Dichloro-1,2,2-trifluoroethane (Freon 123a)	8021	5.0
1,1,2,2-Tetrachloroethene	8021	5.0
Trichlorotrifluoroethane (Freon 113)	8021	5.0
1,1,2-Trichloroethene	8021	5.0
cis-1,2-Dichloroethene	8021	5.0

---

[5] EPA Method SW-846. Method 8010 or 8020 may be substituted for method 8021.

[6] Groundwater protection concentrations for corrective measures.

[7] The sum of 1,2-Dichlorobenzene and 1,4-Dichlorobenzene must be ≤ 4.7 µg/L.

Table C-3. Selected compounds, cont.

Constituents	Method[8]	Conc. Limit (ug/L)[9]
Fluoride[10]	4500-F C[11]	1500.0
Chromium (total)[12]	7191	50.0
Zinc (total)[13]	6010	300.0

---

[8] EPA Method SW-846. Method 8010 or 8020 may be substituted for method 8021.

[9] Groundwater protection concentrations for corrective measures.

[10] Area "A" and Area "C".

[11] APHA, AWWA, WPCF, American Public Health Association. 1989. *Standard Methods for the Examination of Water and Wastewater*. 17th ed., Washington, DC, p. 4-87.

[12] Area "A".

[13] Area "C".

TABLE C-4. Key compounds.

LOCATION	COMPOUND
Area A remediation area	Tetrachloroethene Trichloroethene cis-1,2-Dichloroethene
Area B remediation area	Tetrachloroethene
Area C remediation area	Tetrachloroethene Trichloroethene cis-1,2-Dichloroethene Vinyl Chloride Freon TF
Area D remediation area	Tetrachloroethene Trichloroethene cis-1,2-Dichloroethene
SEQ remediation area	Tetrachloroethene Trichloroethene cis-1,2-Dichloroethene
Area C B/330 investigative area	Tetrachloroethene Trichloroethene cis-1,2-Dichloroethene Vinyl Chloride Freon TF
Building 322 investigative area	Freon TF Freon 123a Tetrachloroethene Trichloroethene cis-1,2-Dichloroethene
Bedrock aquifer	Freon TF Freon 123a Tetrachloroethene Trichloroethene cis-1,2-Dichloroethene

APPENDIX D--REPORTING FORMATS

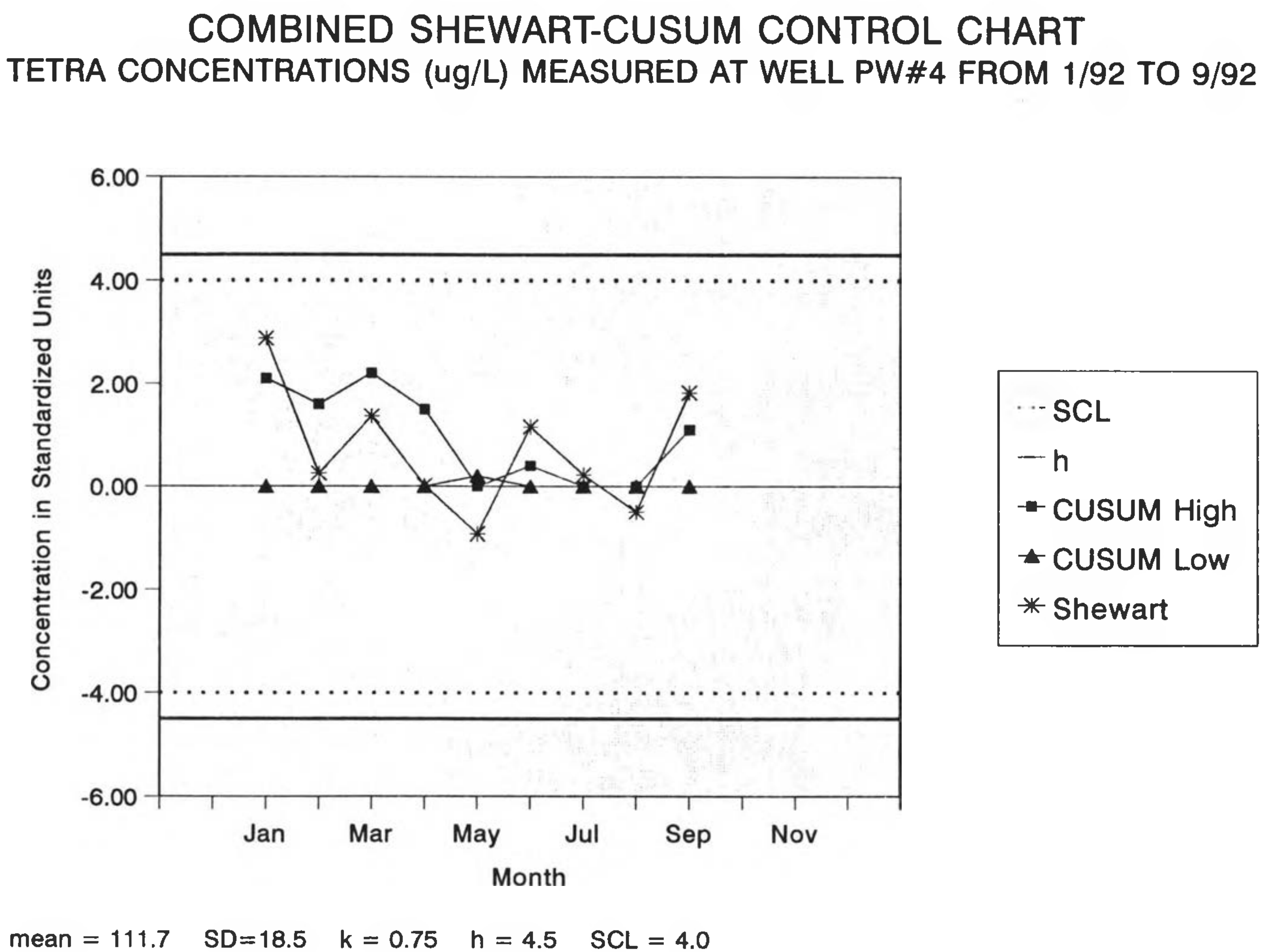


FIGURE D-1. Reporting format for combined Shewart-CUSUM control chart.

Well #	Compound	Group 1		Group 2		$\hat{\Delta}$ (µg/L)	CI% (µg/L)	CL_l (µg/L)	CL_u (µg/L)
		dates	*n*	dates	*m*				
016	CEDC	88-89	4	90-91	4	0.95	89	-0.5	2.4

FIGURE D-2. Reporting format for Hodges-Lehmann estimates and confidence intervals.

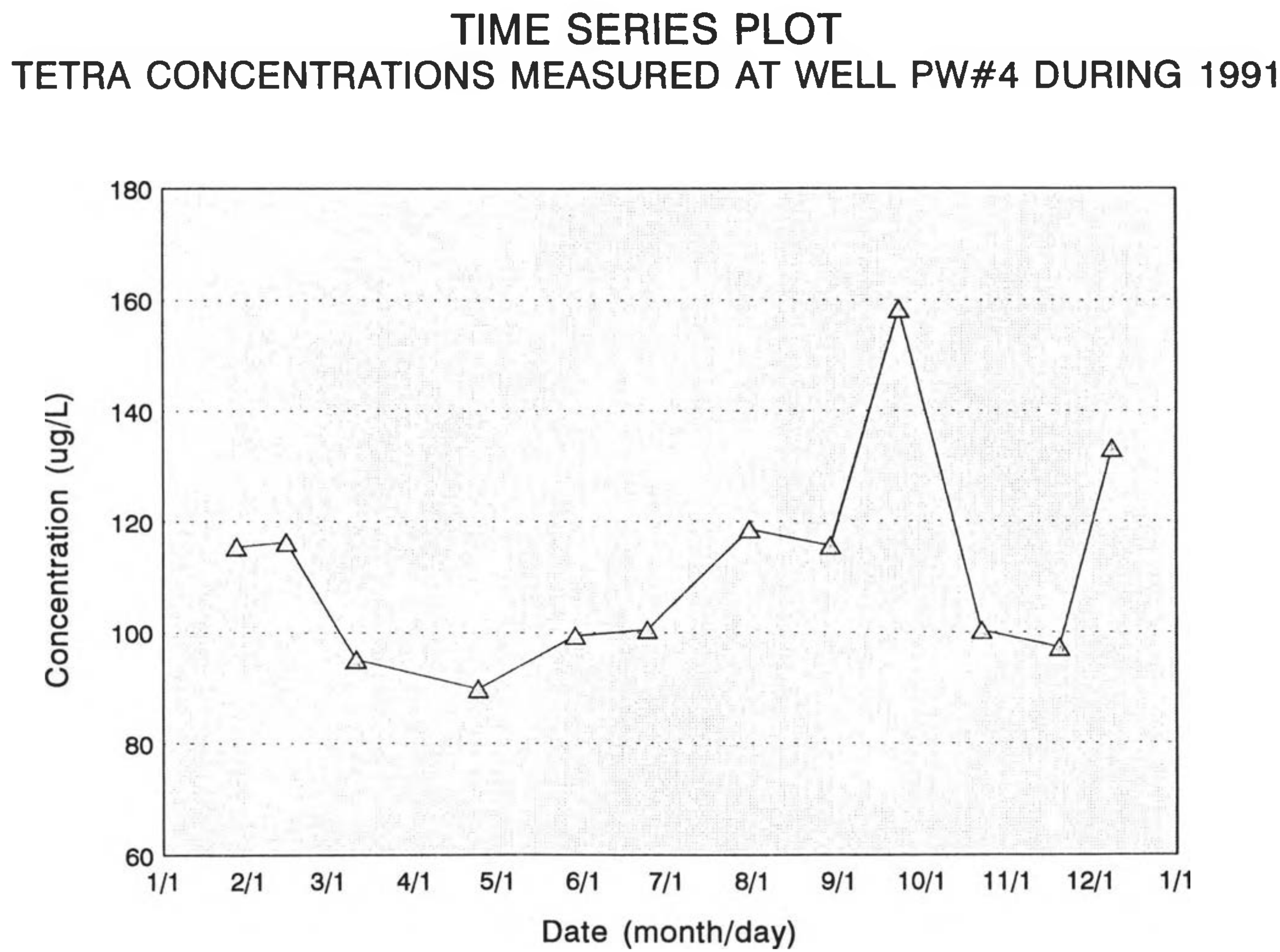


FIGURE D-3. Reporting format for time series plots.

Well #	Compound	Date	n	CI%	CL_l	CL_u
PW#4	TRICE	1986	12	96	2.1	2.4

FIGURE D-4. Reporting format for interval estimates of the median.

APPENDIX E--DATA ANALYSIS METHODS

APPENDIX E-I

COMBINED SHEWART-CUSUM CONTROL CHARTS

INTRODUCTION

The design and implementation of combined Shewart-CUSUM control charts is described in this appendix. Definitions of terms and parameters are presented first, followed by design procedures. The procedures are then used to design the control chart scheme for information goal #2: *to ensure adequacy of the treatment facility by detecting changes in input concentrations*. Implementation procedures are outlined and then applied to data obtained from the IBM East Fishkill facility. Finally, a table of references for control charts is given.

It is assumed that each contaminant at each well will be tracked separately. It is also assumed that, for the most part, only one observation will be obtained per sampling event. Equations are provided, however, that accommodate average values obtained from multiple observations obtained for quality control purposes.

DEFINITIONS OF TERMS

ARL (average run length): The average number of samples before an out-of-control signal is obtained. The ARL should be large when the process is close to the target value and small when the process has shifted too far from the target.

FIR (fast initial response) feature: Assigning a value greater than 0 to the CUSUM "headstart value", S_o.

out-of-control: The situation that exists when the process has shifted too far from the target value.

two-in-a-row rule: A process is declared out-of-control only if two out-of-control signals are obtained in succession.

DEFINITIONS OF PARAMETERS

μ Estimate of the population mean. It is used as the target value for the Shewart-CUSUM quality control scheme.

σ Estimate of the population standard deviation.

SCL Shewart control limit.

k A parameter of the CUSUM scheme known as the "reference value."

h A parameter of the CUSUM scheme known as the "decision interval."

Δ The deviation to be detected by the CUSUM scheme.

S_{Hi} The CUSUM statistic at time = i for positive shifts.

S_{Li} The CUSUM statistic at time = i for negative shifts.

S_o The CUSUM statistic at time = 0. Also known as the "headstart value."

Y_i The ith observation for both the Shewart and CUSUM schemes. It represents a single reading.

$\bar{Y}_i$ The ith observation for both the Shewart and CUSUM schemes. It represents the average from a number of observations.

Z_i The standardized value of the ith observation (or average of observations). It is plotted on the control chart and compared to the *SCL*. It is also used in calculating S_{Hi} and S_{Li}.

DESIGN PROCEDURES

1. **Use the worksheet shown in Figure E-1 to design the control chart**

2. **State the monitoring information goal, monitoring approach and statistical information goal**

3. **Decide whether to use one-sided or two-sided charts**
   Two-sided control charts should be used if both increases and decreases in concentration are of interest.

4. **Choose a value for *SCL***

WORKSHEET FOR DESIGNING SHEWART-CUSUM CONTROL CHARTS

Monitoring Information Goal:

Monitoring Approach:

Statistical Information Goal:

One-sided or Two-sided?

Parameter	Value	How Obtained
SCL		
k		
h		
S_o		

Decision Rule:

Comments:

FIGURE E-1. Worksheet for designing combined Shewart-CUSUM control charts.

5. **Choose a value for *k***

For a given in-control ARL, the quickest detection is obtained when $k = \Delta/2$.

6. **Choose a value for *h***

h can be chosen from a table of average run lengths (ARLs). The value of *h* should be selected to give the desired ARL when the process is both in-control and out-of-control.

7. **Choose a value for S_o**

If the FIR feature is not used, $S_o = 0$. If the FIR feature is used, S_o is frequently set at $h/2$. The FIR feature should be implemented if it is felt that the process is likely to be out-of-control at startup or after a restart following a control action.

8. **Decide whether or not to use the two-in-a-row rule**

The two-in-a-row rule is recommended for situations where outliers are likely. It is particularly suitable if only single observations rather than an average of observations are used, because the effects of an outlier will be smoothed out if averages are taken.

The two outliers do not have to be from the same side of the distribution. Although a suspected outlier on the high side followed by a suspected outlier on the low side may not indicate a shift in the process mean, it would indicate the need for closer examination of water quality.

APPLICATION OF DESIGN PROCEDURES

The above procedures were used to design the control chart for information goal #2. The design is summarized in Figure E-2.

IMPLEMENTATION PROCEDURES

1. **Use the worksheet in Figure E-3 to implement the control chart**

2. **Calculate μ and σ from historical data**

Use data that was collected when the system was operating within desirable limits. Outliers and data that is part of an obvious trend should not be used. Precision of the estimates will increase with increasing sample size.

WORKSHEET FOR DESIGNING SHEWART-CUSUM CONTROL CHARTS

Monitoring Information Goal: To ensure adequacy of the treatment facility by detecting increases in input concentrations.

Monitoring Approach: Detect increases in mean concentration over time that are moderate and persistent or large and sudden.

Statistical Information Goal: Immediately detect concentrations that are $\geq 4\sigma$ above the target mean. Quickly detect persistent changes in concentration that are $\geq 1.5\sigma$ above the target mean.

One-sided or Two-sided? two-sided

Parameter	Value	How Obtained
SCL	4.0	information goal
k	0.75	$k = \Delta/2$: $\Delta = 1.5$ as per information goal
h	4.5	Part 1, Table 2 in Lucas, (1982)
S_o	0.0	process not likely to be out-of-control at startup

Decision Rule: two-in-a-row rule

Comments: The two-in-a-row rule was chosen because outliers are likely.

FIGURE E-2. Design of combined Shewart-CUSUM control chart for information goal #2.

WORKSHEET FOR IMPLEMENTING SHEWART-CUSUM CONTROL CHARTS

Well #: **Compound:**

μ = σ =

CONCENTRATIONS (μg/L) USED FOR CALCULATING μ AND σ					
Date	Conc.	Date	Conc.	Date	Conc.

Month	Y_i or $\bar{Y}_i$ (μg/L)	Z_i	$Z_i - k$	S_{Hi}	$-Z_i - k$	S_{Li}
Jan						
Feb						
Mar						
Apr						
May						
Jun						
Jul						
Aug						
Sep						
Oct						
Nov						
Dec						

Comments:

FIGURE E-3. Worksheet for implementing combined Shewart-CUSUM control charts.

3. **Construct the control chart**
The horizontal axis should be in units of time and the vertical axis in standardized concentration units. Draw in the upper Shewart limit at *SCL*, lower Shewart limit at -*SCL*, upper CUSUM limit at *h* and lower CUSUM limit at -*h*.

4. **Calculate:**

$$Z_i = (Y_i - \mu)/\sigma \text{ for a single observation}$$

$$Z_i = (\bar{Y}_i - \mu)/\sigma \text{ for multiple observations}$$

5. **Calculate:**

$$S_{Hi} = \max\ [0,\ (Z_i - k) + S_{H(i-1)}]$$

$$S_{Li} = \max\ [0,\ (-Z_i - k) + S_{L(i-1)}]$$

6. Plot Z_i, S_{Hi} and $-S_{Li}$ on the control chart.

7. An out-of-control signal is given if $|Z_i| \geq SCL$ or if S_{Hi} or $S_{Li} \geq h$.

8. Declare the process out-of-control if two out-of-control signals occur in succession.

9. If any parameters of the control chart are modified, S_{Hi} and S_{Li} should be reset to 0.

APPLICATION OF IMPLEMENTATION PROCEDURES

The above procedures were used to implement the control chart that was designed for information goal #2. Implementation is summarized in Figure E-4.

UPDATING AND REVISION PROCEDURES

There are three situations where the Shewart-CUSUM control chart may need to be updated or revised.

1. **Recalculate μ and σ after the process is declared out-of-control.**

Control charts were traditionally used for industrial or manufacturing purposes. If a control chart indicated that a process was out-of-control, adjustments were made to the process itself.

The scenario is somewhat different for the information goal stated here because the "process" cannot be modified. Input concentrations are monitored to ensure adequacy of the

WORKSHEET FOR IMPLEMENTING SHEWART-CUSUM CONTROL CHARTS

Well #: PW#4 **Compound:** Tetrachloroethene (TETRA)

μ = 111.7 µg/L σ = 18.5 µg/L

CONCENTRATIONS (µg/L) USED FOR CALCULATING μ AND σ					
Date	**Conc.**	**Date**	**Conc.**	**Date**	**Conc.**
01-28-91	115.5	06-24-91	100.5	11-20-91	97.2
02-15-91	116.3	07-31-91	118.6	12-09-91	133.2
03-11-91	95.1	08-29-91	115.6		
04-24-91	89.8	09-23-91	158.4		
05-29-91	99.4	10-23-91	100.3		

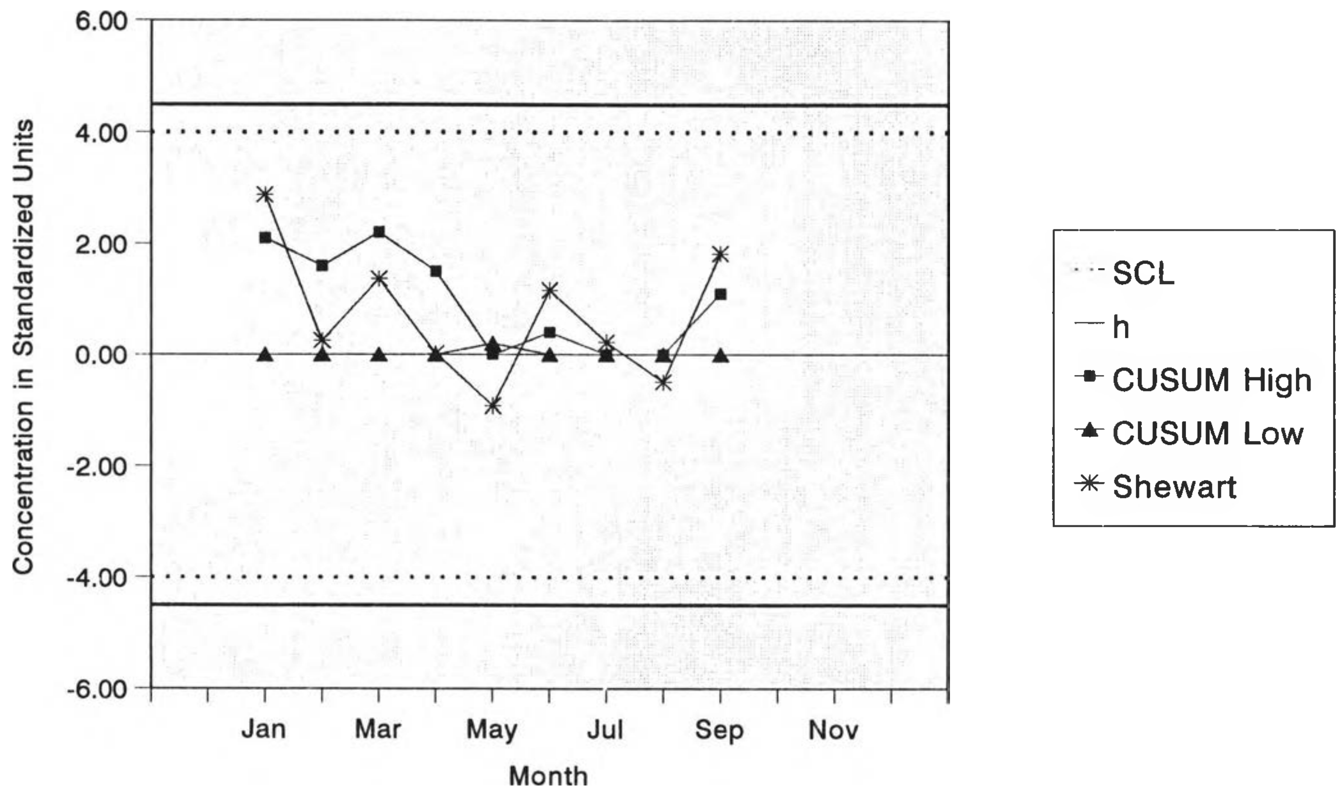
Month	Y_i or $\bar{Y}_i$ (µg/L)	Z_i	$Z_i - k$	S_{Hi}	$-Z_i - k$	S_{Li}
Jan	165.0	2.88	2.13	2.1	-3.63	0
Feb	116.4	0.25	-0.50	1.6	-1.00	0
Mar	137.0	1.36	0.61	2.2	-2.11	0
Apr	112.0	0.02	-0.73	1.5	-0.77	0
May	94.6	-0.92	-1.67	0	0.17	0.2
Jun	133.1	1.16	0.41	0.4	-1.91	0
Jul	115.8	0.22	-0.53	0	-0.97	0
Aug	102.5	-0.50	-1.25	0	-0.25	0
Sep	145.4	1.82	1.07	1.1	-2.57	0
Oct						
Nov						
Dec						

Comments: Z_i, S_{Hi} and $-S_{Li}$ are plotted on the control chart shown in Figure E-5. No out-of-control signals occurred.

FIGURE E-4. Implementation of combined Shewart-CUSUM control chart for information goal #2.

COMBINED SHEWART-CUSUM CONTROL CHART

TETRA CONCENTRATIONS (ug/L) MEASURED AT WELL PW#4 FROM 1/92 TO 9/92



mean = 111.7   SD=18.5   k = 0.75   h = 4.5   SCL = 4.0

FIGURE E-5. Plot of Shewart-CUSUM data from Figure E-4.

treatment system. If the control chart indicates that the process is out-of-control, the treatment system or perhaps the flow are adjusted, not the input concentrations.

For example, if the control chart indicates a persistent increase in TETRA concentrations, the treatment system may have to be modified to handle the larger concentrations. At that point it might be appropriate to revise μ (and maybe σ) prior to restarting the control chart.

For the information goal listed here, μ and σ will be examined and, if necessary, recalculated after each out-of-control episode.

2. **Improve the precision of μ and σ as more data becomes available.**

New data should periodically be added to the original data that was used to calculate μ and σ. The parameters can then be updated. For the information goal listed here, μ and σ will be updated annually.

3. **Revise control chart parameters to better serve information needs.**

There are many ways to revise Shewart-CUSUM control charts to meet information needs. For example, if the control chart is too sensitive, values of SCL, h and/or k can be increased. This type of revision will be conducted according to section VII of this data analysis protocol.

REFERENCES

Reference	Comments
Berthouex and Hunter, 1975	Discusses the use of Shewart charts for monitoring sewage treatment plants.
Berthouex *et al.*, 1978	Examines the effect of data record attributes on the use of Shewart and CUSUM charts for monitoring sewage treatment plants.
Bissell, 1984	Evaluates the performance of control charts in the presence of serial correlation.
Cheremisinoff, 1988	Describes how to construct control charts using Lotus 123.
Hockman and Lucas, 1987	A well-written explanation of CUSUM schemes. Also describes how to use MINITAB to construct CUSUM charts.
Lucas, 1982	An excellent article on combined Shewart-CUSUM control charts.
Lucas, 1985	Discusses the FIR feature and the two-in-a-row rule.
Lucas and Crosier, 1982	Discusses the two-in-a-row rule.
Page, 1954 and 1961	Describes the development of the CUSUM scheme.
Schilling and Nelson, 1976	Examines the effect of nonnormality on the control limits of Shewart charts.
Starks, 1989	Evaluates the use control chart methods for RCRA sites.
Starks and Flatman, 1991	Basically the same as Starks, 1989.
U.S. EPA, 1989	Section 7 is devoted to the use of control charts at RCRA facilities.
Vasilopoulos and Stamboulis, 1978	Describes how to modify control chart limits in the presence of data correlation.
Vaughan and Russell, 1983	Discusses the use of control charts to monitor point source discharge.

APPENDIX E-II

THE HODGES-LEHMANN ESTIMATOR AND CONFIDENCE INTERVAL

INTRODUCTION

The Hodges-Lehmann estimator, $\hat{\Delta}$, is a nonparametric estimate of the difference between two independent groups. $\hat{\Delta}$ is the median of all possible pairwise differences between the x and y values.

DEFINITIONS OF PARAMETERS

$\hat{\Delta}$ Hodges-Lehmann estimator: $\hat{\Delta}$ = median $[x_i - y_i]$ for x_i, $i = 1, \ldots n$ and $j = 1, \ldots m$.

x_i The ith observation of group 1.

y_i The ith observation of group 2.

n The number of observations in group 1.

m The number of observations in group 2.

N The number of pairwise differences: $N = n \times m$.

x^* The critical value from a table for the rank sum test. Used to calculate R_l and R_u for small sample sizes.

$Z_{\alpha/2}$ The critical value from a table of standard normal quantiles. Used to calculate R_l and R_u for large sample sizes.

R_l The rank of the pairwise difference that is the lower confidence limit of $\hat{\Delta}$.

R_u The rank of the pairwise difference that is the upper confidence limit of $\hat{\Delta}$.

CL_l The lower confidence limit of $\hat{\Delta}$.

CL_u The upper confidence limit of $\hat{\Delta}$.

INFORMATION GOALS

Monitoring Information Goal: To determine the lateral extent of contamination plumes.

Statistical Approach: Compare recent concentration measurements to older measurements to see how much change there has been in contaminant concentration.

Statistical Information Goal: Estimate the magnitude of the difference between the current two years of concentration data and the previous two years of concentration data. Determine a 90 percent confidence interval for the estimate.

CALCULATION PROCEDURES

1. Calculate all pairwise differences.

2. Rank the pairwise differences from smallest to largest.

3. Calculate the median of the pairwise differences. This is the Hodges-Lehmann estimator.

4. If $m + n \leq 20$, go to step 5. Otherwise, go to step 7.

---

5. Refer to a table of quantiles for the rank sum statistic. Find the critical value x^* nearest to $\alpha/2$.

6. **Calculate:**

$$R_l = x^* - \frac{n(n+1)}{2}$$

$$R_u = N - R_l + 1$$

---

7. Refer to a table of standard normal quantiles. Find $Z_{\alpha/2}$.

8. **Calculate:**

$$R_l = \frac{N - Z_{\alpha/2}\sqrt{\frac{N(n+m+1)}{3}}}{2}$$

$$R_u = N - R_l + 1$$

---

9. Find the pairwise differences that correspond to R_l and R_u. These values are CL_l and CL_u, the upper and lower bounds of the confidence interval. The true difference will lie between CL_l and CL_u an average of (100-α) percent of the time.

EXAMPLE

Well #: 016

Compound: cis-1,2-Dichloroethene (CEDC)

Group 1		**Group 2**	
date	**conc. (μg/L)**	**date**	**conc. (μg/L)**
03-03-88	7.9	03-01-90	5.4
09-08-88	6.4	09-06-90	5.8
03-08-89	6.8	03-07-91	5.5
09-14-89	5.5	09-09-91	6.9

Pairwise Differences		Ordered Pairwise Differences	
2.5	1.4	-1.4	1.0
2.1	1.0	-0.5	1.0
2.4	1.3	-0.3	1.0
1.0	-0.1	-0.1	1.3
1.0	0.1	0.0	1.4
0.6	-0.3	0.1	2.1
0.9	0.0	0.6	2.4
-0.5	-1.4	0.9	2.5

median: (0.9 + 1.0)/2 = 0.95 µg/L

x^* = 12 (for α = 0.114)

R_l = 2 R_u = 15

CL_l = -0.5 µg/L CL_u = 2.4 µg/L

$\hat{\Delta}$ = 0.95 with an 89 percent confidence interval of -0.5 to 2.4

REFERENCES

Reference	Comments
Helsel and Hirsch, 1992	Explains the Hodges-Lehmann estimator and confidence interval. Includes tables (Table B4).
Hirsch, 1988	Examines robustness of the Hodges-Lehmann estimator. Develops the seasonal Hodges-Lehmann estimator.
Hodges and Lehmann, 1963	Presents the original development of the Hodges-Lehmann estimator.
Hollander and Wolfe, 1973	Explains the Hodges-Lehmann estimator and confidence interval. Includes tables (Table A.5).

APPENDIX E-III

INTERVAL ESTIMATE OF THE MEDIAN

INTRODUCTION

A nonparametric interval estimate of the true population median is described here. The estimate is calculated using the binomial distribution.

DEFINITIONS OF PARAMETERS

x_i The *i*th observation of the sample.

n The number of observations in the sample.

x' The critical value from a table for the sign test (or a binomial table). Used to calculate R_l and R_u for small sample sizes.

$Z_{\alpha/2}$ The critical value from a table of standard normal quantiles. Used to calculate R_l and R_u for large sample sizes.

R_l The rank of the observation that is the lower confidence limit of $C_{0.5}$.

R_u The rank of the observation that is the upper confidence limit of $C_{0.5}$.

$C_{0.5}$ The true population median.

CL_l The lower confidence limit of $C_{0.5}$.

CL_u The upper confidence limit of $C_{0.5}$.

INFORMATION GOALS

Monitoring Information Goal: To measure water quality down-gradient of AOCs where flow may pass across a site boundary.

Statistical Approach: Determine the range of concentrations that will contain the true population median a large percent of the time.

Statistical Information Goal: On an annual basis, estimate the 95 percent confidence interval of the true population median concentration.

CALCULATION PROCEDURES

1. Rank the sample observations from smallest to largest.

2. If $n \leq 20$, go to step 3. Otherwise, go to step 5.

---

5. Refer to a table of quantiles for the sign test statistic. (If a binomial table is used, enter it at the p = 0.5 column). Find the critical value x' nearest to $\alpha/2$.

6. **Calculate:**

$$R_l = x' + 1$$

$$R_u = n - x' = x$$

---

7. Refer to a table of standard normal quantiles. Find $Z_{\alpha/2}$.

8. **Calculate:**

$$R_l = \frac{n - Z_{\alpha/2}\sqrt{n}}{2}$$

$$R_u = \frac{n + Z_{\alpha/2}\sqrt{n}}{2} + 1$$

---

9. Find the observation that correspond to R_l and R_u. These values are CL_l and CL_u, the upper and lower bounds of the confidence interval. The true population median will lie between CL_l and CL_u an average of (100-α) percent of the time.

10. Declare the concentrations out of compliance if the entire confidence interval lies above the GWPS (see Figure E-6).

EXAMPLE

Well #: PW#4

Compound: Trichloroethene (TRICE)

Observations		Observations	
date	**conc. (µg/L)**	**date**	**conc. (µg/L)**
01-13-86	2.4	07-02-86	2.4
02-19-86	2.6	08-06-86	2.2
03-11-86	2.1	09-25-86	2.6
04-15-86	2.0	10-27-86	2.4
05-14-86	2.3	11-24-86	2.4
06-19-86	2.2	12-16-86	2.1

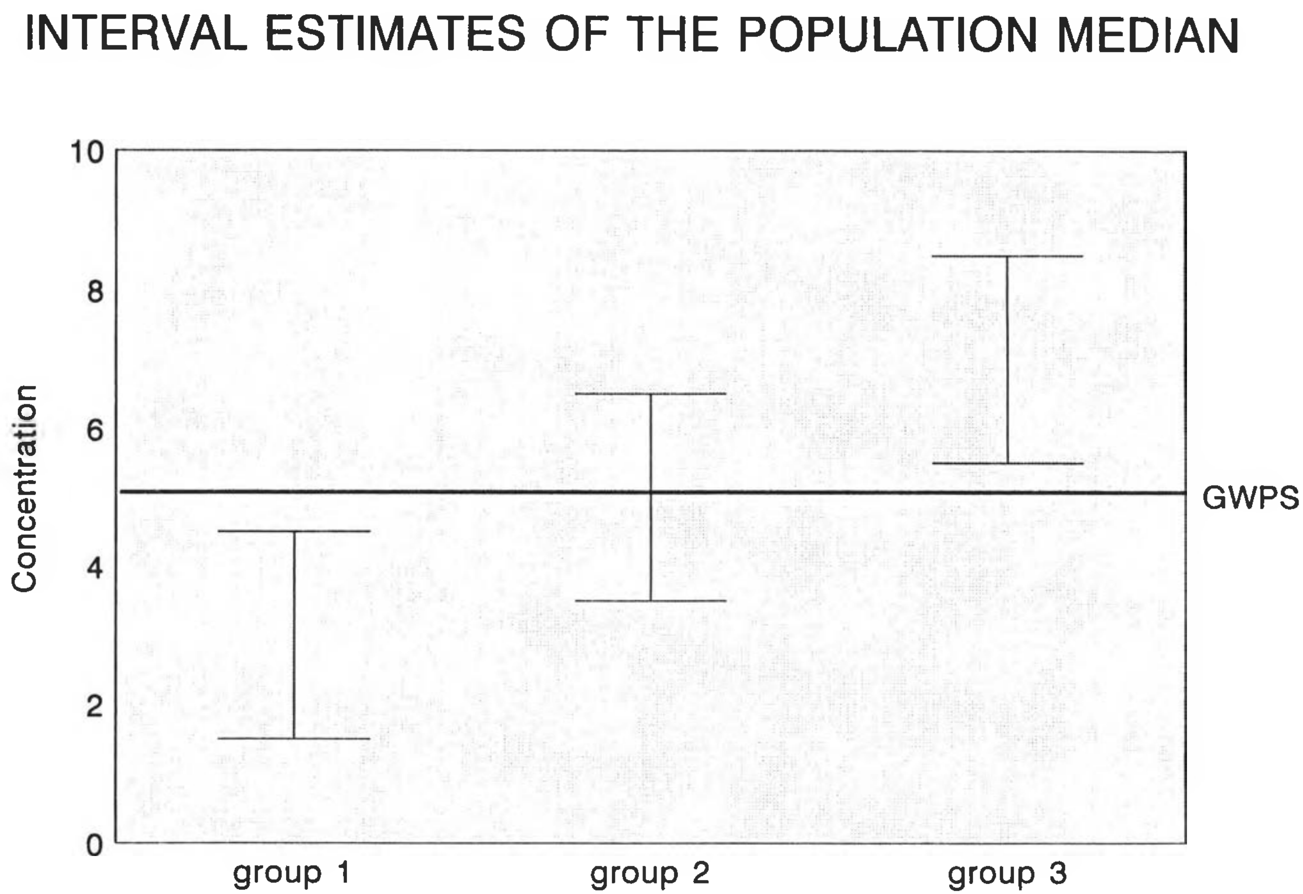
Ordered Observations	
2.0	2.4
2.1	2.4
2.1	2.4
2.2	2.4
2.2	2.6
2.3	2.6

$x' = 2$ (for $\alpha = 0.0193$)

$R_l = 3$ $R_u = 10$

$CL_l = 2.1$ µg/L $CL_u = 2.4$ µg/L

The 96 percent confidence interval of the true population median is 2.1 to 2.4.



Groups 1 and 2 are in compliance.
Group 3 is out of compliance.

FIGURE E-6. Comparison of confidence intervals to the groundwater protection standard (GWPS).

REFERENCES

Reference	Comments
Helsel and Hirsch, 1992	Explains how to obtain the nonparametric interval estimate of the median. Includes tables (Table B5).