Methods for determining compliance with groundwater quality regulations at waste disposal facilities : final report. [DNR-014b] 1989

Fisher, Sarah R.; Potter, Kenneth W.

Madison, Wisconsin: Wisconsin Department of Natural Resources, 1989

Wisconsin Groundwater Management Practice Monitoring Project No. 16

GROUNDWATER
Wisconsin's
buried treasure

Wisconsin Department of Natural Resources

Methods for Determining Compliance with Groundwater Quality Regulations at Waste Disposal Facilities

FINAL REPORT

January 1989

Submitted to: **Wisconsin Department of Natural Resources**
Bureau of Water Resources Management and
Bureau of Solid and Hazardous Waste Management

Submitted by: **Sarah R. Fisher and Kenneth W. Potter**
Department of Civil and Environmental Engineering
University of Wisconsin - Madison

Acknowledgements

I would like to acknowledge those people whose time and effort made this report possible. First I would like to thank my mother and father because without their support and encouragement I could not have pursued a career as an environmental engineer.

The environmental issues addressed in this work are multidimensional with political, scientific and social aspects. Yet, the day to day grind was primarily computer related. My advisor, Professor Ken Potter, provided invaluable consulting on all aspects of the problem. I thank him for always having an open door. Thanks are also due to Professor Peter Bosscher who helped with computer data base management and data transfer problems. Mr. John Murer helped to evaluate contaminant hydrogeology at many facilities. Mr. Pete Schoemann also deserves many thank-you's for technical and editorial review, for generating figures, and for making dinner when I was working too hard.

This work was funded in part by a grant from the Wisconsin Department of Natural Resources (DNR). I am indepted to Mr. Dave Linsdorff and Mr. Jack Connelly of the DNR for making this research possible. I would also like to acknowledge Mr. Steve Alas for helping me obtain DNR file information and for technical review.

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ANOVA	One way analysis of variance
DNR	Wisconsin Department of Natural Resources
EPA	United States Environmental Protection Agency
ES	Enforcement standard
IQR	Interquartile range (fourth spread)
KW	Kruskal-Wallis test
HWS	Hazardous waste site
MSWLF	Municipal solid waste landfill
PAL	Preventive action limit
RCRA	Resource Conservation and Recovery Act

CHAPTER 1
INTRODUCTION

1.0 Overview

The groundwater resources in the State of Wisconsin are protected by the Wisconsin Department of Natural Resources (DNR). Groundwater quality standards are established under Chapter NR 140 of the Wisconsin Administrative Code (Wisconsin DNR, 1988).. Also addressed in NR 140 are statistical methods for 1) evaluating background (clean) water quality and 2) determining exceedance of a water quality standard or finding an environmentally significant change in water quality. In this paper we address the complex issue of how to "defensibly" establish background water quality at waste disposal facilities, and subsequently how to "defensibly" discern standard exceedances and/or significant water quality changes. "Defensibly" is meant to imply that the technical approach should be acceptable to concerned parties and, if necessary, in a court of law. In the regulatory context, the intent of this study is to evaluate alternative analytic methods to meet the objectives of NR 140. Analytic techniques are also evaluated with respect to existing and proposed federal regulations for hazardous waste facilities and for municipal solid waste landfills. While the focus of this paper is on solid waste disposal facilities, the techniques are also applicable to most types of hazardous waste sites, land disposal systems and storage facilities.

To help the DNR prioritize its work in enforcing NR 140, a secondary goal of this research is to screen the licensed landfill sites in Wisconsin for evidence of contamination. While the analytic methods discussed above will help to define the degree and extent of contamination at each site, the statistical screening will help steer the DNR towards particular sites. More specific objectives of this study include:

- to summarize current information on the statistical properties of groundwater quality data, and to evaluate how these properties affect the choice of analytic method;
- to investigate the hydrogeology and water quality at 20 landfill sites in order to document defensible procedures for establishing background water quality and determining apparent contamination;
- to detail procedures for establishing background water quality;

- to evaluate appropriate statistical tests for determination of significant changes in water quality; and,
- to investigate site-wide "predictors" of groundwater contamination to prioritize licensed sites for regulatory action.

This report is the culmination of a two-part study which began in 1985. The first report (Goodman and Potter, 1987) is entitled "Graphical and Statistical Methods to Assess the Effect of Landfills on Groundwater Quality." The earlier study focused on 1) evaluating the statistical nature of groundwater quality data and 2) developing procedures for evaluating groundwater quality. These procedures are currently being used by the DNR. In this report, the earlier results are summarized and in some instances methods are expanded or modified based on additional research.

This report is written assuming that the reader has a rudimentary understanding of probability and statistics as well as hydrogeology, contaminant transport and water chemistry. Other readers will be able to understand the basic concepts presented. For the reader interested primarily in regulatory issues and recommendations, the chapter summaries and all of Chapter 5 should be sufficient.

The report is organized as follows:

- Chapter 1 includes an overview of state and federal regulations and places them in a statistical context. Also the DNR landfill water quality database is introduced including a summary of the 20 landfill sites studied in detail in this study.
- Graphical techniques for visualization of water quality data are presented in Chapter 2. Also, the statistical properties of groundwater quality data are evaluated. These properties are introduced in the context of the geophysical environment and related contaminant transport processes.
- Chapter 3 addresses how to evaluate groundwater contamination, given the statistical properties of the data. Types of statistical tests are introduced in Section 3.0. Statistical tests are evaluated with respect to basic assumptions, performance and utility in Sections 3.1. 3.2 and 3.3. The applicability of Sections 3.1 to 3.3 to existing and proposed regulations are discussed in Sections 3.4 and 3.5.
- A predictor of groundwater quality change is introduced in Chapter 4. This predictor is used to execute a statistical screening of the groundwater

quality database. The use of the predictor for prioritizing regulatory work is discussed.

- Conclusions and recommendations are summarized in Chapter 5. Also, flow charts are used to present simple procedures to 1) define background water quality and 2) determine compliance of waste sites with existing regulations.

1.1 Statistical Context of Regulations

The DNR and EPA regulations addressed in the following sections recognize that groundwater quality data vary temporally and spatially due to natural effects, and are also affected by sampling and analytic error[1]. Due to natural variability, the determination of a change in water quality should be linked to probability theory. Two regulatory situations present themselves:

1) Has a water quality standard been exceeded?

2) Has there been a significant change in water quality?

The first question is the more straight forward. If a sample value exceeds a standard, accounting for sampling and analytic variability, then a violation has occurred. In this situation a "violation" means only that a mandated concentration level has been exceeded, not that certain actions must be taken. Defining the magnitude of sampling and analytic variability is addressed in Chapter 3. The general idea is that a standard may be exceeded a "little" due to laboratory error, before a "defensible" violation occurs. This problem is acute when standards are at or approach the level of detection of the contaminant, as is the case with some volatile organic compounds. Sampling error may be addressed by timely resampling of the entire site.

The second question is more complex, since now a comparison must be made between supposedly "clean" background data and possibly contaminated data, both of which are subject to temporal and spatial variability as well as sampling and analytic error. Hence,

[1]Sampling error in this context refers to error introduced by the technician during well sampling. Analytic error occurs in the laboratory. These are not to be confused with natural "sampling" error associated with spatial and temporal fluctuations in water quality.

the problem becomes one of statistical inference. The question could be rephrased as a statistical hypothesis:

Null Hypothesis: H_0: No Contamination exists; facility is in compliance

Alternative Hypothesis: H_1: Contamination exists; facility is in violation

A statistical test is made on the null hypothesis and a conclusion is reached that either the facility is in violation or the facility is not. In this situation, a "violation" implies that water quality is significantly different from background, not that certain actions must be taken. The conclusion is based on probability assessment. Figure 1-1 illustrates the two types of errors associated with hypothesis testing.

TRUE SITUATION	STATISTICAL DECISION: IN COMPLIANCE	STATISTICAL DECISION: IN VIOLATION
IN COMPLIANCE	Good Decision $1-\alpha$	False Positive Decision α Type I error
IN VIOLATIION	False Negative Decision β Type II Error	Good Decision $1-\beta$ "Power" of Test

FIGURE 1-1 Statistical Error in Hypothesis Testing

Type I error (false positive decision) occurs when a site (or well) is actually in compliance but the statistical test determines it to be in violation. The probability of a Type I error is defined as the significance level of the test, α, and can be controlled. Usually, α is set at 0.05, giving a 1/20 chance that a "false positive" conclusion of contamination will occur. α is exact however only when the assumptions of the test are met. The Type I error rate sets the "level of protection" afforded the polluter since when a Type I error occurs a site owner may be required to perform remedial measures when none are necessary.

On the other hand Type II error (false negative decision) occurs when contamination exists but is not detected. The Type II error sets the "level of protection" afforded the State (i.e. the people and the environment). Unfortunately, the probability of a false negative conclusion, b, is not controlled, is often difficult to calculate, and is dependent on many factors which may include sample size, magnitude of "change" in concentration, and choice of statistical test. Because the Type II error rate is usually unknown and is likely to be higher than the Type I error rate, the ability for the State to protect the environment is confounded with the ability to maintain a low false negative rate.

Hypothesis tests may be divided into two general categories: those which rely on the estimation of parameters of a probability distribution (usually the mean and standard deviation of the normal distribution) and those which do not. The former are generally referred to as parametric procedures, while the latter are nonparametric. Sometimes nonparametric tests are referred to as distribution-free methods, although this name may be misleading (See Section 3.0). Examples of parametric procedures are Student's t-test developed by W.S. Gossett in 1908, and the commonly-used analysis of variance. Nonparametric methods usually rely on test statistics developed from the ordered ranks of the data. For example, when testing for an increasing trend in time, if the concentration data are ranked in exact order of time, "perfect" correlation would be found. The most widely known nonparametric measure is the median, or middle value of a data set.

These fundamental concepts are introduced here to place the state and federal regulations in context with the statistical hypothesis problem.

1.2 Wisconsin Regulatory Context

Wisconsin has established two types of water quality standards: enforcement standards (ES) and preventive action limits (PAL's). ES's are maximum contaminant levels and are not to be exceeded. They are set forth in NR 140.10 for public health related substances and in NR 140.12 for public welfare related substances (Wisconsin DNR, 1988). For these substances, the PAL is established as a percentage of the ES. The PAL in this situation is a "flag" of potential contamination. 60 substances have mandated ES's and PAL's. If natural water quality exceeds the mandated ES or PAL, an alternative concentration limit (ACL) may be set by the DNR.

PAL's, for substances without an ES, are defined based on background water quality. Substances with PAL's set based on background water quality are called "indicator parameters." An indicator parameter is a naturally occurring substance which is indicative of groundwater degradation when high values are observed or when significant concentration changes occur. High concentration of a "true" indicator, is not necessarily a health or welfare problem. Examples are specific conductance, total hardness, pH and alkalinity. Under the current regulations, a PAL for an indicator parameter is set by calculating the mean and standard deviation of background water quality (based on a minimum of 8 representative data points). The PAL is then set as :

$$PAL = \bar{x} + 3\,s$$

where $\bar{x}$ is the sample mean and s is the sample standard deviation. In some cases s is so small that the difference between the PAL and background water quality is not environmentally significant. In such cases the PAL is based on a table of minimum significant increases above the background mean. PAL's set based on minimum increases defined in Table 3 of NR 140 are called "table values" in this report.

The choice of the mean plus three standard deviations as a measure of environmental significance is based in part on a DNR study of 16 landfills, where it was found that clean well concentrations rarely exceeded $\bar{x} + 3\,s$ of background water quality. Statistically, this method implies that there is less than a one percent chance that a truly "clean" data point will exceed the PAL, if the data are independent and normally distributed about a mean value. These statistical assumptions may or may not be true as discussed in detail in Chapter 2. Hence, in this situation, it may be best to view a PAL as an empirical estimate of environmentally significant change. If water quality data are assumed to be stationary, independent, and normally distributed, then, in the case of no contamination, the probability that a PAL would be exceeded by a single measurement would be about 1 percent.

In this study, we have focused on eight chemical parameters as shown in Table 1-1. These substances are the most frequently monitored parameters at landfill sites in Wisconsin. Specific conductance, total alkalinity, total hardness, pH, and chemical oxygen demand are considered "indicator parameters." Iron, chloride, and sulfate are welfare concerns and, therefore, have enforcement standards as shown.

The Wisconsin regulations require the DNR to use a scientifically valid statistical procedure to determine if an enforcement standard is attained or exceeded or if a change in the concentration of a substance has occurred. In NR 140.14 (2) the following statistical procedures are specified at a 95% level of confidence:

(a) Student t-test;

(b) Temporal or spatial trend analysis; or

(c) Other scientifically valid statistical analyses which are appropriate for the data being considered.

The main objective of this study is to provide technical guidance to DNR on the implementation of these or other procedures for determining compliance with NR 140. These regulations are similar to EPA's existing and proposed regulations for determining compliance at hazardous waste and municipal solid waste facilities which are discussed in the next section.

TABLE 1-1
Chemical parameters included in analyses and applicable groundwater quality standards.

PARAMETER	ENFORCEMENT STANDARD (mg/l)	PREVENTIVE ACTION LIMIT (mg/l)
Total Alkalinity as $CaCO_3$	----	*
Total Hardness	----	*
pH[1]	----	*
Specific Conductance[1,4]	----	*
Chemical Oxygen Demand	----	*
Iron - Total[2]	0.30[3]	0.15[3]
Chloride	250[3]	125[3]
Sulfate	250[3]	125[3]

1 Field measurements only
2 Total includes dissolved and suspended particulate material
3 If background water quality exhibits high concentrations an ACL may be established.
4 Units are µmhos/cm
* PAL set based on background water quality.

1.3 Federal Regulatory Context

Like the DNR, the EPA has wrestled with statistical evaluation of groundwater quality data at landfills. The EPA also has established water quality standards for specific chemical parameters The State of Wisconsin standards, discussed in the previous section are as strict, or stricter, than current federal standards.

Federal regulations also exist for statistical determination of compliance for RCRA facilities. These are existing and new hazardous waste facilities covered by Subtitle C of the Resource Conservation and Recovery Act (RCRA) and regulated by 40 CFR Parts 264 and 265. Until recently, Part 264 Subpart F provided that Cochran's Approximation to the Behrens Fisher Student's t-test (CABF) or an alternative statistical procedure approved by EPA be used to determine whether there is a statistically significant exceedance of background levels, or other allowable levels, of specified constituents. These regulations, and in particular the CABF procedure, generated criticism and EPA proposed a new regulation in response to these concerns (EPA, August 24, 1987). The proposed regulation was revised based on comments EPA received and was then made final (EPA, October 11, 1988). A draft guidance document for implementation of these regulations is currently under final review (EPA, 1988).

The final regulation establishes five performance standards that a statistical procedure must meet. The performance standards allow flexibility in designing statistical procedures to site specific considerations. The choice of an appropriate statistical test can be made based on the data available, the hydrogeology of the site and the theoretical properties of the test. The proposed regulations do recommend four types of statistical procedures to detect contamination in groundwater. For more information see Section 3.1.

In addition EPA is currently drafting amendments to Subtitle D of RCRA to include criteria for municipal solid waste landfills (MSWLF's). The statistical test requirements are similar to the RCRA Subtitle C final regulation and recommend the same four types of procedures. The big difference between the federal hazardous waste and solid waste regulations is that at the permitted hazardous waste facilities, four <u>independent</u> samples must be collected for each monitoring round (the sampling interval may be as large as monthly to obtain independence (EPA, October 11,1988)). At MSWLF's only one quarterly or semi-annual measurement is necessary.

Unlike the Wisconsin rules the proposed regulations do not incorporate the idea of "exceeding background values or concentration limits" in terms of "a minimum increase." According to the Federal regulation, this is because any statistically significant increase is a cause for concern.

Another important difference between the EPA and Wisconsin regulations is that the EPA regulations require that the statistical tests be applied quarterly to new data. The tests provide a tool to determine which of two phases of groundwater monitoring is necessary. Under Subtitle C these phases are "detection monitoring" and the more extensive "compliance monitoring." The draft Subtitle D regulations divide monitoring into "Phase I" and "Phase II." Thus, the statistics are used as a gate into stricter regulatory control (i.e. more extensive monitoring and possible remediation). In Wisconsin the statistical tests are not a quarterly requirement; the statistics are only a possible tool to either confirm a standard exceedance or to detect a change in water quality. They do not have any direct consequences associated with them. In this respect, the draft Subtitle D regulations are stricter than current Wisconsin rules for non-hazardous waste disposal facilities.

EPA performance standards and recommended procedures for statistical analysis of groundwater quality data are addressed in detail in Chapter 3. In order to evaluate the statistical tests, we investigated 20 Wisconsin landfill sites. An overview of these sites is presented in the next section.

1.4 Wisconsin Landfill Water Quality Database

The Wisconsin DNR Bureau of Solid and Hazardous Waste Management has collected groundwater quality data at solid waste disposal sites for many years. The landfill groundwater database includes over 300 licensed sites, each with a number of wells, and water quality data for an array of constituents at each well. The majority of sites are typically older, unengineered sites. Many are now closed or have clay-lined expansions. The database also includes many county-owned state-of-the-art sanitary landfills, as well as some older county-run facilities. Several sites in the database are considered to have seriously contaminated groundwater with hazardous substances.

Also included are industrial sites owned by paper mills, electric utilities and a variety of other industries.

We obtained water quality and water elevation data for 316 licensed landfill facilities from the DNR in August, 1987. The results reported in Chapter 4, for contamination predictors, are based on analysis of this database. In April, 1988 updated data were obtained for 20 sites chosen for detailed analysis (Chapters 2 and 3). Due to the time lag for laboratory analysis, data transmittal and computer entry, these data may be considered current at least through the end of 1987.

The locations of the 20 selected sites are shown on Figure 1-2 and a summary of site characteristics is presented in Table 1-2. Nine (9) of the 20 were included in the previous DNR-funded study (Goodman and Potter, 1987). The sites may be generally classified by ownership and design as follows:

9 small to medium size municipally-owned solid waste sites, either unlined or partially unlined with an engineered expansion;

3 industrial facilities:
  2 paper sludge,
  1 fly ash;

3 larger "natural attenuation"[2] or unlined county-owned solid waste sites, and;

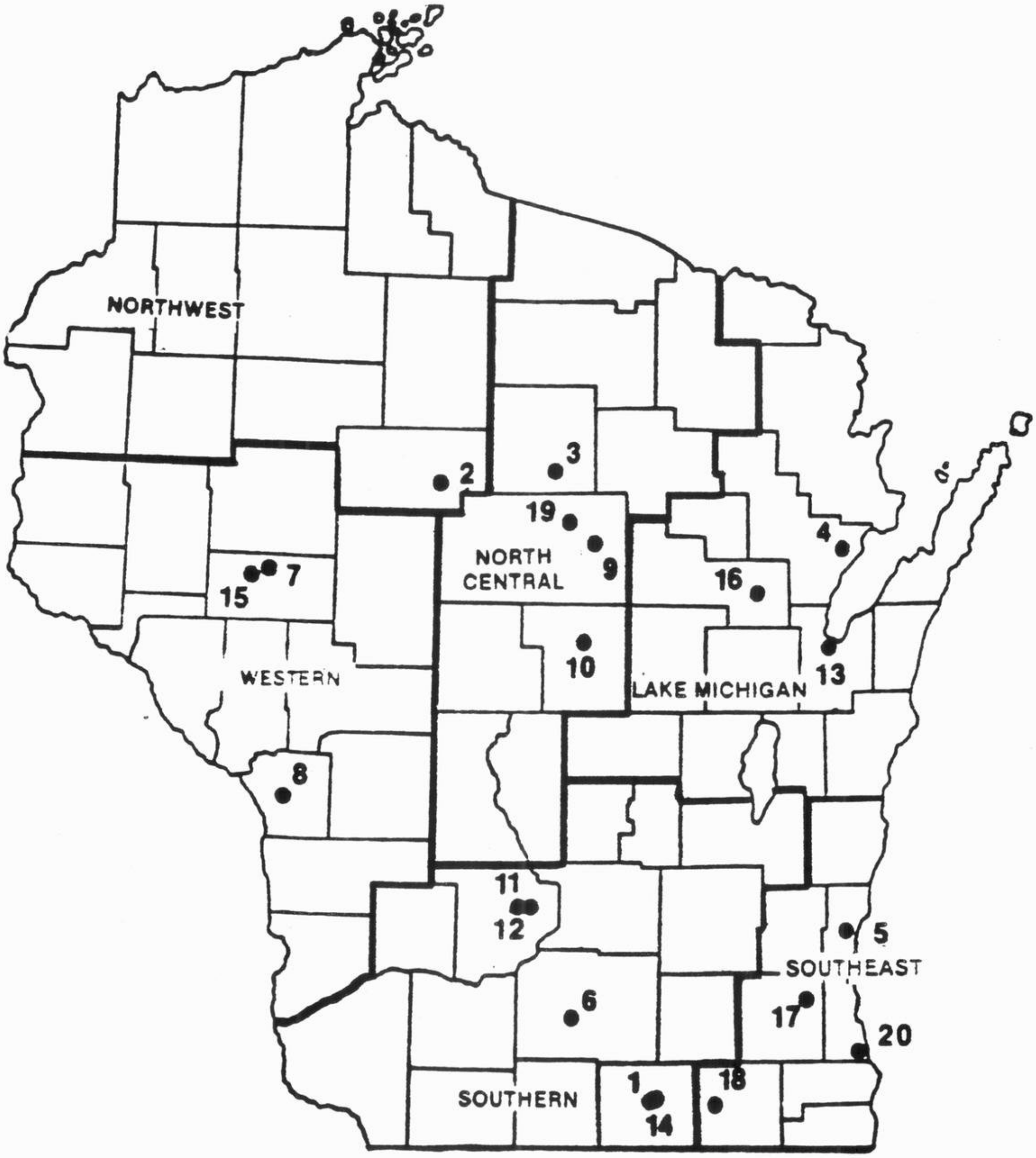
5 county-owned clay lined solid waste facilities with leachate collection systems

---

20 total

2 "Natural attenuation" implies that the uppermost subsoil is a dense silt or clay material which should minimize, but not necessarily eliminate, leachate movement.

THE STATE OF WISCONSIN WITH COUNTY BOUNDARIES
AND
DNR FIELD DISTRICTS



1	City of Janesville (old)	11	Sauk County (old)
2	City of Medford	12	Sauk County (new)
3	City of Merrill	13	Fort Howard Paper Co.
4	City of Oconto	14	Rock County/City ofJanesville
5	City/Town of Cedarburg	15	Town of Washington
6	Dane County - Verona	16	Village of Bonduel
7	Eau Claire County - 7 Mile Creek	17	WMI Brookfield
8	La Crosse County	18	WMI Greidanus
9	Marathon County	19	Wausau Paper Mills
10	Portage County	20	WEPCO Oak Creek Ash Disposal

FIGURE 1-2 Name and location of facilities studied

TABLE 1-2: Characterization of landfill sites studied

FACILITY NAME (LICENSE)	FACILITY DESIGN	WASTE RECEIVED	DESIGN VOLUME $yd^3 \cdot 10^6$	SITE SIZE (Acres)	SITE LIFE	YEAR SAMPLING BEGAN AND (Avg. No. of data pts)	NUMBER OF WELLS
City of Janesville (2822)	Partial clay lined, partial unlined	MSW, IND	0.7	18	1961 - 1978	1982 (19)	11
City of Medford (341)	Unlined	MSW	NONE	4.5	1972 - ?	1980 - 83 (10)* 1983 (19)	6 OLD 6 NEW
City of Merrill (912)	Unlined	MSW	NONE			1975 (40)	18
City of Oconto (137)	Natural attenuation, 7/83 groundwater interception trench installed.	MSW, IND	NONE	14.5	Early '70's - 83	1977 (36)	14
City/Town of Cedarburg (271)	Natural attenuation	MSW	NONE	10	1972 - 87	1975 (41)	11
County Dane #1 - Verona (2680)	Natural attenuation Partial leachate collection system	MSW	2	49	1977 - Pres.	1977 (41)	23
County Eau Claire Seven Mile Creek (2821)	Clay lined, leachate collection	MSW	1.2	24	1978 - Pres.	1978 ( )	8
County Lacrosse (2637)	Natural attenuation	MSW, IND	1.38	55	1976 - Pres.	1977 ( )	12
County Marathon (2892)	Clay lined, leachate collection	MSW, IND	1.5	10	1980 - Pres.	1980 ( )	11
County Portage (2966)	Clay lined, leachate collection	MSW	1.28	18.6	1984 - Pres.	1984 (20)	19

* Two different sets of wells: old and new

TABLE 1-2 (continued)

FACILITY NAME (LICENSE)	FACILITY DESIGN	WASTE RECEIVED	DESIGN VOLUME $yd^3 \cdot 10^6$	SITE SIZE (Acres)	SITE LIFE	YEAR SAMPLING BEGAN AND (Avg. No. of data pts)	NUMBER OF WELLS
County Sauk (Old) (2051)	NONE	MSW foundary sand	1.0	4	1973 - 83	1979	16
County Sauk (New) (2978)	Clay lined, leachate collection	MSW	1.28	20	1983 - Pres.	1983 (20)	12
Fort Howard Paper Co. Green Bay (2332)	Part unlined, part lined	Paper sludge	4.5	293	1964 - Pres.		23
Rock County City of Janesville (3023)	Clay lined, leachate collection	MSW		45	1985 - Pres.	1984	16
Town of Washington (160)	Unlined	MSW and Unknown	NONE	6.5	1930's - 79	1977 (35)	6
Village of Bonduel (59)	Unlined	MSW, Canning waste and Pickles	NONE	<5	1950's - Pres.	1986 (9)	6
Waste Management Inc. Greidanus Landfill (140)	Unlined, Capped 1988	MSW	NONE	33	1969 - 1981	1976 (47)	15
Wsste Management Inc. Brookfield Landfill (1)	Part Nat. atten, part clay lined	MSW and unknown	OLD: NONE NEW:	O: ( ) N: ( )	1970 - Pres. Part closed	1977 (37)	16
Wausau Paper Mills (2875)	Mostly clay lined leachate collection	Paper sludge	0.25	4.8	1981 - Pres.	1982 (21)	8
Wisconsin Electric Power -- OakCreek (2357)	Unlined	Fly ash	4.0	130	1975 - Pres.	1975 ( ) SA	17

The 20 sites range in size from 4 acres to 293 acres. The newest site opened in 1983, while the oldest has existed since the 1930's. Groundwater monitoring did not become widespread until the mid to late 1970's as evidenced by the years of water quality record. Three of the sites are either listed or proposed to be included on the EPA's national priority list for hazardous waste cleanup funding. The number of monitoring wells listed reflects the number of wells with sufficient water quality data for statistical analysis (greater than 8 sampling dates for several parameters). A total of 274 wells were included in the analysis.

The 20 landfills are all located in the lower two-thirds of the State. Bedrock in the eastern edge of the state is primarily dolomite underlaid by limestone and/or shale (6/20 sites). A large part of the state is covered with sand and gravel and underlaid by sandstone (7 sites). The southwest portion of Wisconsin was unglaciated in the most recent "Wisconsin" glaciation of Pleistocene age (5/20 sites). Four sites located in the north lie in an area with a thin unconsolidated zone over Precambrian age middle and lower proterozoic rocks (granite, etc).

Each of Wisconsin's four major hydrogeologic provinces are represented. Zaporozec and Cotter (1985) define hydrogeologic conditions in these provinces and in nine subdistricts in terms of Pleistocene unconsolidated deposits. To some extent, the hydrogeologic provinces mimic the bedrock geology.

The 20 selected sites represent a cross-section of Wisconsin solid waste disposal facilities. The industrial sector may be under-represented with only 3/20 sites. While industrial waste streams may be quite different from municipal solid waste, this does not imply that different analytic methods are needed to detect industrial waste groundwater contamination. The big issue is the same. Is this facility contaminating groundwater, and if so, what is the degree and extent of contamination?

CHAPTER TWO
GROUNDWATER QUALITY DATA

2.0 Overview

In this chapter the statistical nature of groundwater quality data will be explored. More specifically, in Section 2-1 graphical methods are introduced and some basic statistical concepts are reviewed. In Section 2.2 the assumptions implicit to statistical hypothesis testing are reviewed. Groundwater quality data are evaluated with respect to the validity or violation of these assumptions.

Natural groundwater quality is known to vary both spatially -- between wells -- and temporally -- at a single well. Anthropogenic effects also contribute to the variability observed in water quality data. In order to understand the specifics of groundwater contamination at a site, the sources of natural variability should be understood and the impact of human activities considered. Sources of variability and error in groundwater data are listed in Table 2-1. A quick glance at this long list illustrates in general the complexity of the problem. It is no wonder that there is much debate in the literature as to appropriate analytic methods.

Natural spatial variability is often due to variations in lithology within the aquifer (Sen, 1982). In Wisconsin many landfills are in areas with glacial till, a poorly sorted soil composed of mixed minerals and rock types (Sugden and John, 1984). Soil and rock heterogeneity may cause the chemical composition of groundwater to vary even at short distances. Spatial variation in water quality data may be exacerbated by well installation and development methods, as well as sampling techniques (Doctor et al, 1985a).

Temporal variability is most often attributed to hydrologic processes. Seasonal effects are usually associated with annual cycles in precipitation and recharge events, particularly for shallow, unconfined aquifers and in areas where surface water/aquifer interactions are significant (Harris *et al*, 1987). Also seasonal pumping for irrigation and high summer input from non-point pollution sources may be causes for seasonal fluctuations in background water quality (Doctor et al, 1985a). Seasonal variation has been reported by several investigators. A literature review on seasonality in groundwater data is presented by Montgomery et al (1987).

TABLE 2-1 Sources of Variability and Human Error in Groundwater Quality Data (Adapted from Doctor *et al*, 1985)

SPATIAL	TEMPORAL	WELL CONSTRUCTION AND DEVELOPMENT	SAMPLE COLLECTION AND ANALYSIS
GEOLOGIC PROPERTIES - lithologic composition, sorting and grain size - structure of lithologic units - bedding planes - fractures (joints and faults) - soil development - properties of vadose zone HYDRAULIC CONDITIONS - Location of recharge/ discharge zones - proximity of water - presence of aquitards - pumping OTHER - other chemical sources - non-point source inputs	TRENDS SEASONAL - recharge - irrigation - fertilization - pesticide/herbicide application - frozen ground PERIODIC - short term precipitation - pumping - river flooding	DRILLING PROCESS - drilling fluids - type of borehole - inter-aquifer transport of materials WELL DESIGN - casing and screen material - diameter - screen length,depth, slot size - packing material - annular seal WELL DEVELOPMENT	COLLECTION - purging method - purging rate/duration - sampling apparatus - cross-contamination between wells - field versus laboratory measurements - sample preparation filtering/container/ preservatives/storage time - operator error - incomplete well development ANALYTIC ERROR - analytic methods, apparatus - operator experience - instrument calibration - interference from other constituents - holding time - clerical errors

The relative importance of these sources of variability is clearly site specific. In general however it is safe to say that natural temporal and spatial variability are greater in magnitude than sampling and analytic error, unless gross sample contamination or mishandling occurs (Doctor et al, 1985a). Groundwater quality, in the local area of a waste facility, appears to vary temporally more than spatially -- as shown in Chapter 3. This may not be true on a regional basis or in some geologic and climatic situations (Sen,1987).

2.1 Visualization

Graphical display of groundwater data is essential. Typically a first step in evaluating groundwater quality is to review existing hydrogeologic information and to try to define groundwater flow and subsurface stratigraphy. The next logical step is to graph the chemical data as concentration versus time. Contaminant "plumes" in plan view or cross-section could also be prepared. Figure 2-1 shows a site map for the City of Merrill landfill, a small municipal unlined facility located above sand and gravel between the confluence of two streams. Groundwater flow is south. The site has 18 wells, 6 of which are shown. Of the six, OB-13 and OB-6 are up or side gradient and are not within the hydraulic influence of groundwater flowing beneath the site. Data for specific conductance are plotted versus time on Figure 2-2. This plot clearly shows increasing trends in time and high relative concentrations for wells OB-2, OB-11, OB-10 and OB-17. It is clear that these points are affected by the landfill. A statistical summary of these data is presented in Table 2-2 with important terms defined.

The same data are presented in box and whisker plots (or "box plots") on Figure 2-3. These plots separate each well and show clearly the difference in the distributions of the data. These plots are generated by ranking the data and may be constructed in different ways (McGill et al, 1978). In this report a software program called STATVIEW 512+ was used. In this program box plots are made as shown on Figure 2-4(a). At the DNR box plots are currently made using software called STATGRAPHICS and are defined as shown in Figure 2-4(b). The boxes are constructed using the median (middle value of the data) and the interquartile range (the range of the middle fifty percent of the data). Note that the median and interquartile range (IQR) are analogous to the more common mean and standard deviation of a set of data. The mean and median are measures of "central tendency" or "location", whereas the standard deviation and IQR are measures of

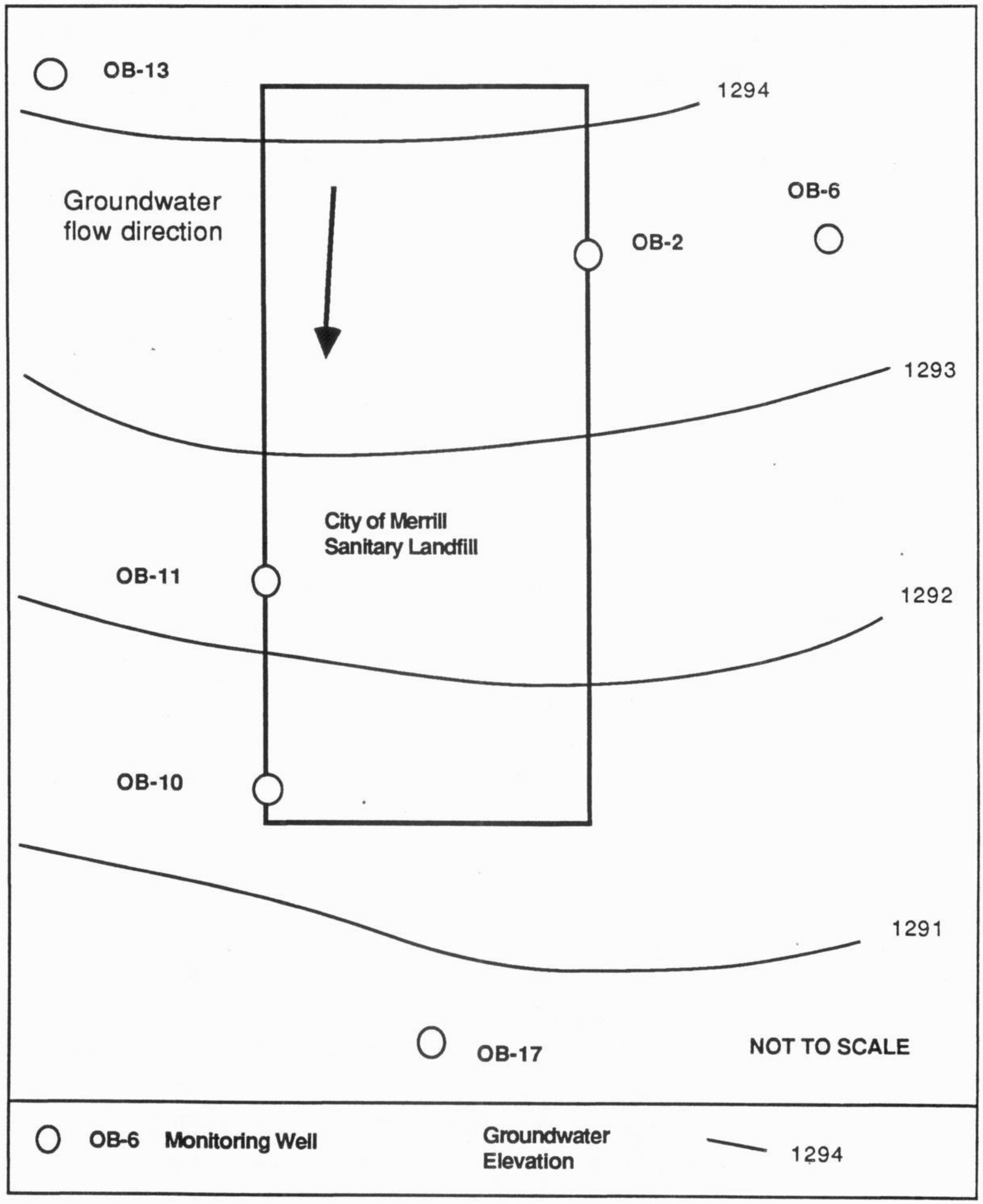


FIGURE 2 - 1 City of Merrill landfill map with selected monitoring wells shown.

TABLE 2-2 Summary Statistics for Specific Conductance with terms defined.

WELL POINT	n	MEAN	MEDIAN	STANDARD DEVIATION	INTER-QUARTILE RANGE	SKEWNESS COEFFICIENT
OB - 13	14	225.3	221.5	36.92	31	.903
OB - 6	33	139.9	117	110.78	47.5	3.796
OB - 2	33	689.1	700	145.28	178.75	-.738
OB - 11	30	851.1	767.5	600.62	1044	.534
OB - 10	31	1321.6	1250	832.06	1199	.397
OB - 17	10	666.0	778.5	246.89	431	-.562

n = Sample size

x_i denotes the raw data

$x[j]$ denotes the order statistics (or ordered ranks) of the data

Median

The middle value of a ranked data set

if n odd: m = x[j], where j= n+1/2
if n even: m = (x[k] +x[k +1]) /2
where k= n/2

Interquartile Range

The range of the middle fifty percent of the data.

Mean =

$$\bar{x} = \frac{1}{n}\sum_{i=}^{n} (x_i)$$

Standard Deviation =

$$s = \sqrt{\frac{1}{n-1}\left(\sum_{i=}^{n} (x_i - \bar{x})^2\right)}$$

Skewness Coefficient =

$$\hat{\gamma} = \frac{\sqrt{n}\left(\sum_{i=1}^{n} (x_i - \bar{x})^3\right)}{\sum_{i=1}^{n} (x_i - \bar{x})^{3/2}}$$

FIGURE 2-2 TIME vs CONCENTRATION PLOT

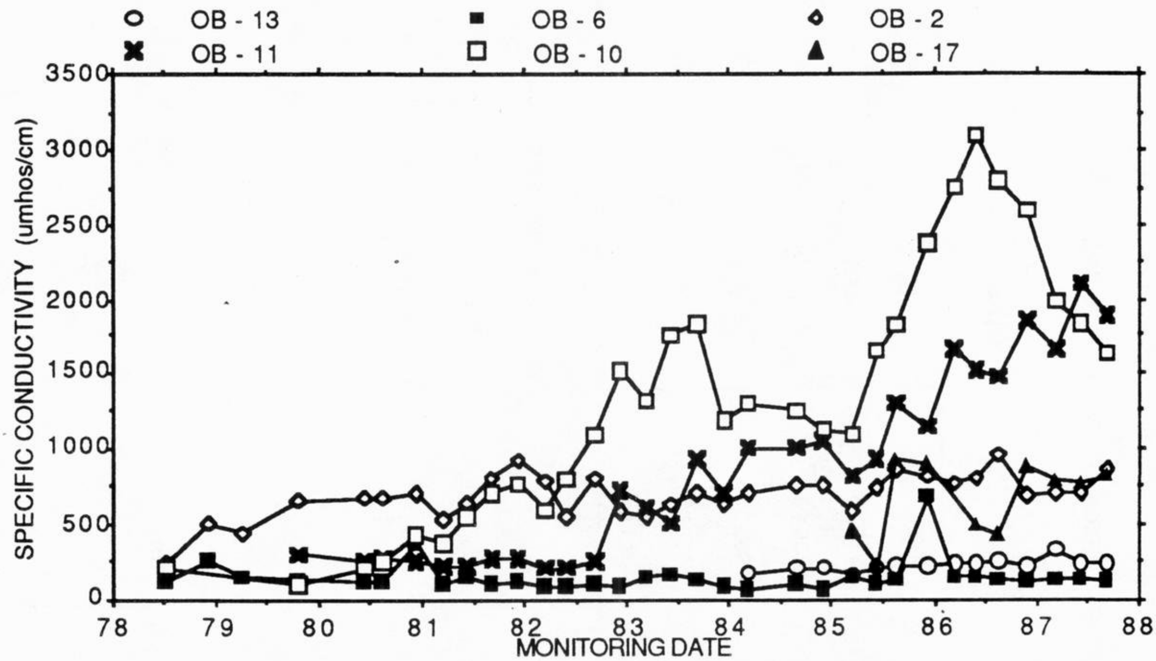
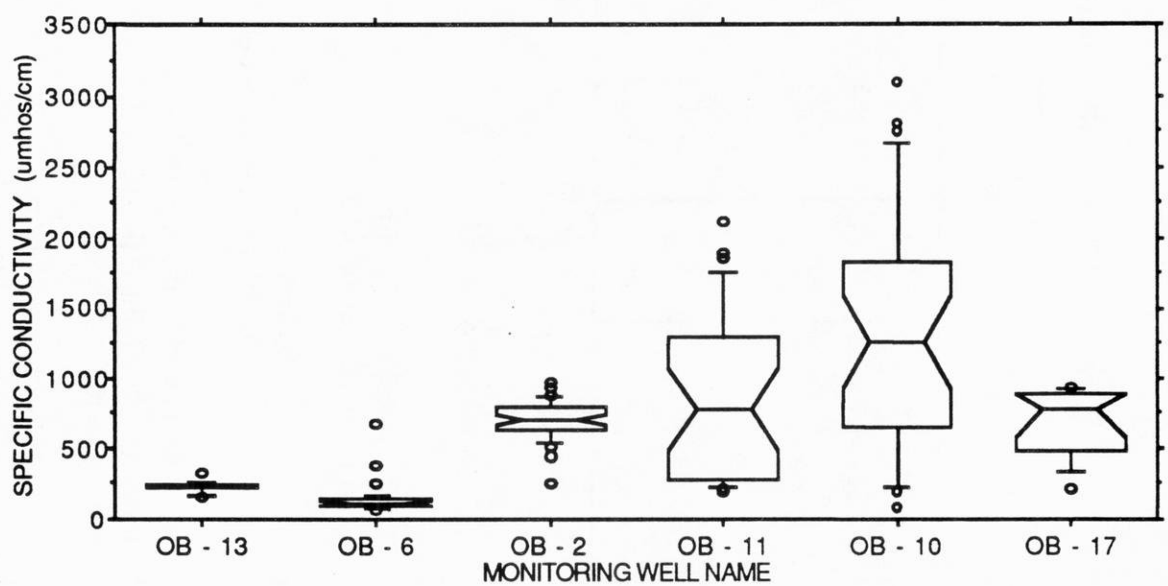


FIGURE 2-3 BOX AND WHISKER PLOTS



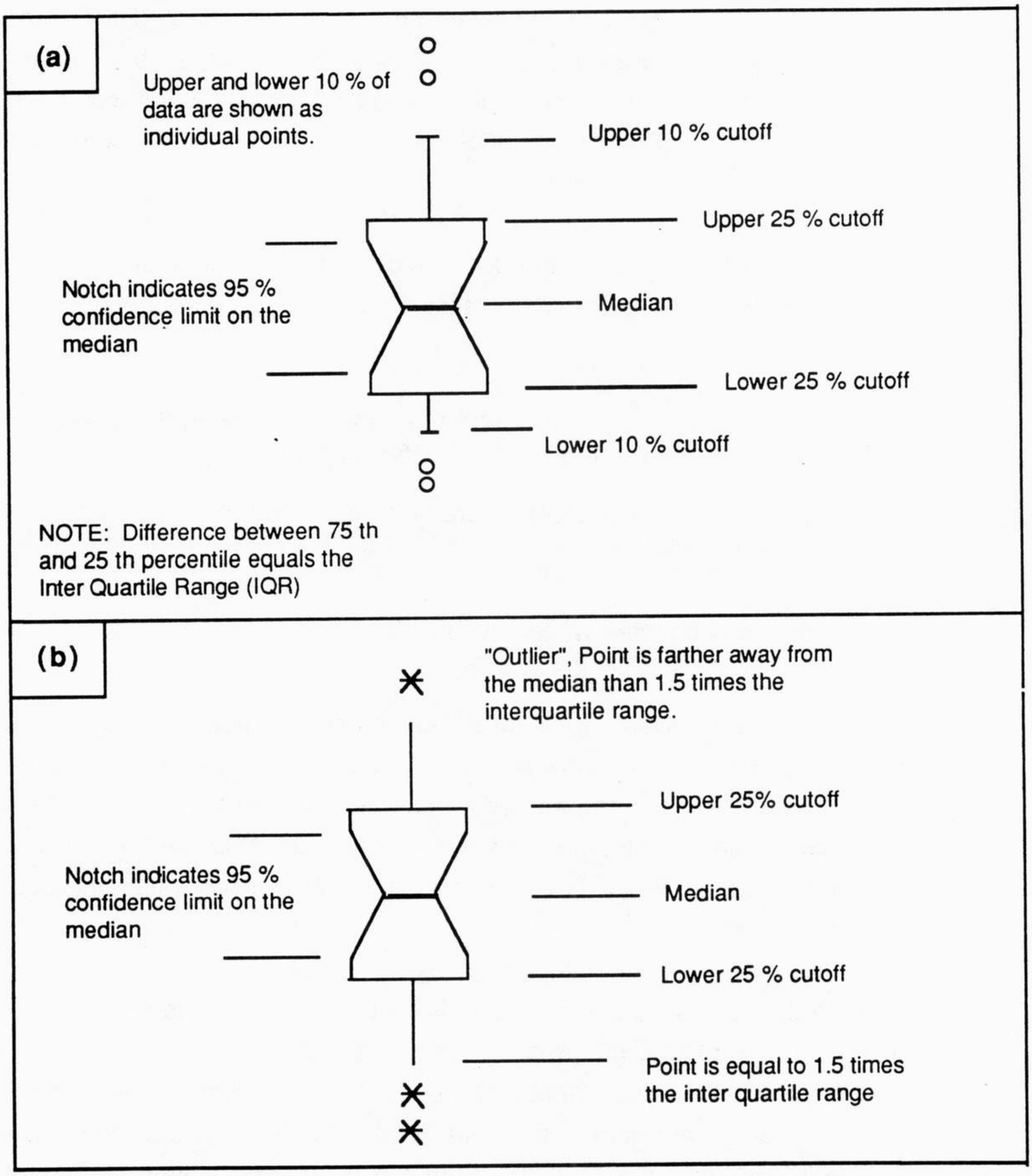


FIGURE 2-4: Box plot construction by STATVIEW 512 + (a) and STATGRAPHICS (b).

"variability." Comparing the two software packages, STATGRAPHICS' boxplots are more informative. STATVIEW always defines outliers as the outer 20 percent of the distribution. It may be that only the outer 5 percent of the data are outliers or it may be that none of the data are outliers. In STATGRAPHICS outliers are defined more explicitly (See Figure 2-3).

Considering again Figures 2-2 and 2-3 it is clear that the box plots also illustrate the apparent contamination at the Merrill site. The following points may be made from Figure 2-3.

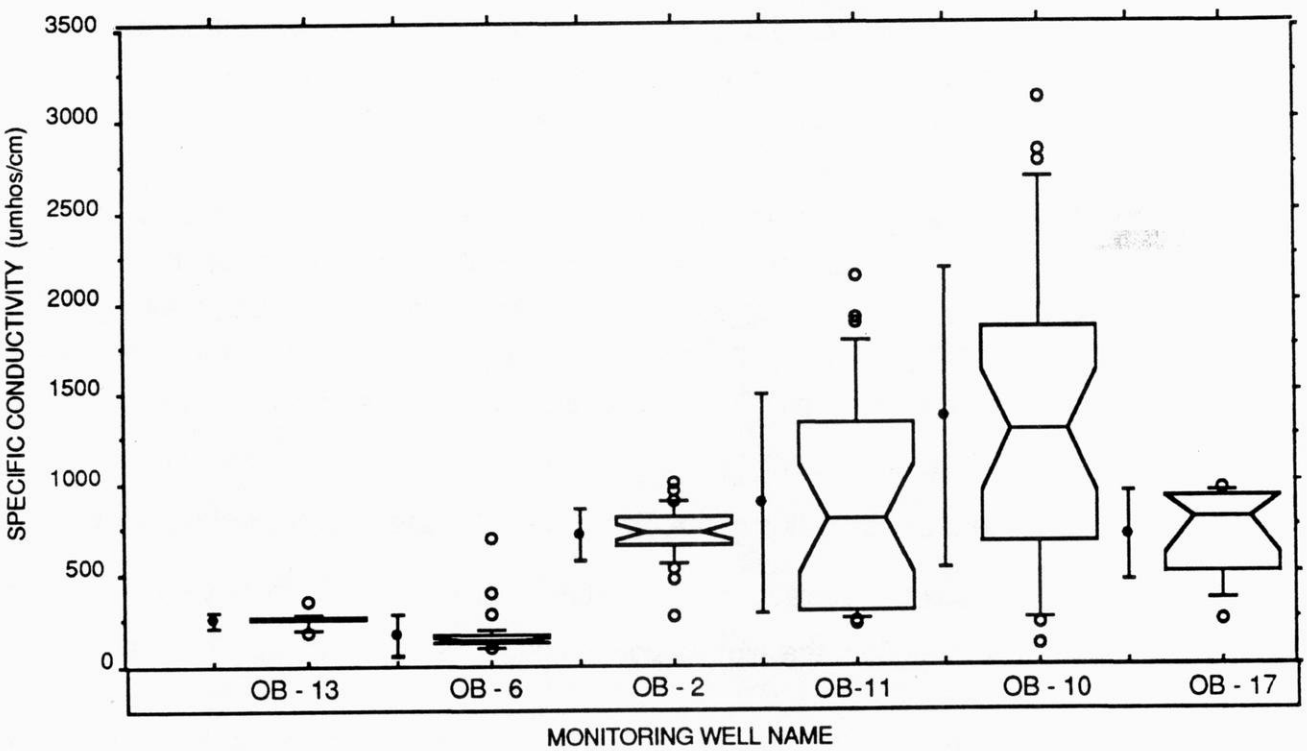
- The two clean wells OB-13 and OB-6 show slightly different background water quality, indicative of natural spatial variability in water quality.
- Wells OB-2 and OB-17 are less impacted than Wells OB-10 and OB-11, indicating that the most intense contamination is directly beneath the landfill as one might expect.
- The impacted wells have much higher variability in the data as evidenced by the wider boxes.

Box plots are very powerful tools for evaluating contamination. At a more complicated site they may be used to even more advantage. For example, all wells screened in similar stratigraphic unit may be aggregated on one plot, or data from two or three well "nests" may be plotted on one plot to illustrate vertical trends. Also water elevation data could be plotted to get a preliminary view of upgradient/downgradient relationships. Many possibilities exist.

The same data are plotted a third time on Figure 2-5. This chart shows the mean values (solid circle) and plus or minus one standard deviation error bars (vertical line) for each well next to each box plot. Note that the mean is consistently greater than the median and two standard deviations is larger than the IQR. <u>This is because high values -- outliers -- tend to inflate the estimate of the mean and standard deviation.</u> The median and IQR, because they are based on ranks, are not sensitive to outlying values. Similar to Figure 2-3, the high variability in the impacted data is shown by the wide error bars.

The box plots are more powerful in visualizing contamination than the error bar plots because they contain more information about the actual distribution of the data. The error bar plots, however, may be useful when working with parametric statistics; for example, when setting PAL's for indicator parameters. For example, if a PAL was calculated from well OB-6 as $\bar{x}_{OB-6} + 3\, s_{OB-6}$, exceedances would be found in all wells except well OB-13. By measuring the three standard deviation point from OB-6, you can see that the highest outlier at this well would be considered an exceedance.

FIGURE 2-5 ONE STANDARD DEVIATION ERROR BAR PLOT



The last graphical displays to be introduced here are the histogram and normal probability plot (also known as the quantile plot or Q-Q plot). Figure 2-6 (a),(b),(c) and (d) show histograms, a tool which may be used to investigate the probability distribution of the data. In simplest terms, the higher the bar the greater probability that (new) measurements will fall in this range. The more sample values the histogram is made from the closer the graph is to the "true" population distribution. Many statistical tests rely on the assumption that the data are drawn from a normal distribution. A comparison of the data distribution to that of the normal may be used to qualitatively evaluate the validity of this assumption. Figure 2-6 (a) and (b) show histograms of conductivity data for Merrill wells OB-6 and OB-11 (note that the scales are different). For comparison Figures 2-6 (c) and (d) are histograms of the normal and lognormal distribution[1]. 1000 variates were generated to construct these figures.

As Figure 2-6 shows neither of Merrill wells OB-6 and OB-11 appear to have normally distributed data; both sets of data are "skewed" to the right. Skew may be conceptualized easily by considering a histogram as a weight and beam balance. The balancing point is the mean. If the data are not symmetric about the mean, but have a long right tail the distribution is said to be positively skewed. The lognormal distribution is also skewed right as shown. Often a transformation of positively skewed data to natural log scale will make the data appear more normal (See Section 2.2).

While the histogram is useful to visualize the probability distribution of the data, it is not the best way to graphically compare data to the normal distribution. Normal probability plots are as easy to construct and give a better representation of the data (Benjamin and Cornell, 1970). Figure 2-7 shows a normal probability plot for the same data as in Figure 2-6 (a) from Merrill well OB-11. These are constructed by first ordering the raw data from smallest to largest. Let $x_{[1]} < x_{[2]} < \ldots < x_{[n]}$ denote the ordered data. The $x_{[i]}$ are called the order statistics of the data. The $x_{[i]}$ are then plotted on normal probability paper versus the corresponding plotting position of $\left(\frac{i}{n+1}\right) \times 100$. If the data are from a normal distribution, the plotted points should lie approximately on a straight line[2].

As one can see in Figure 2-7 (a) the data do not appear to plot as a straight line and we may conclude that the assumption of normality is suspect. If we transform the data to log scale and replot the data as shown on Figure 2-7 (b) the line is not really any straighter, and we cannot conclude that the lognormal distribution is more appropriate. For a full discussion on tests for normality see Section 2.2.1.

1 The log normal distribution may be transformed to the normal by taking the natural log of each variate.

2 The plotting position represents the approximate cumulative probability of a measurement being less than the value observed.

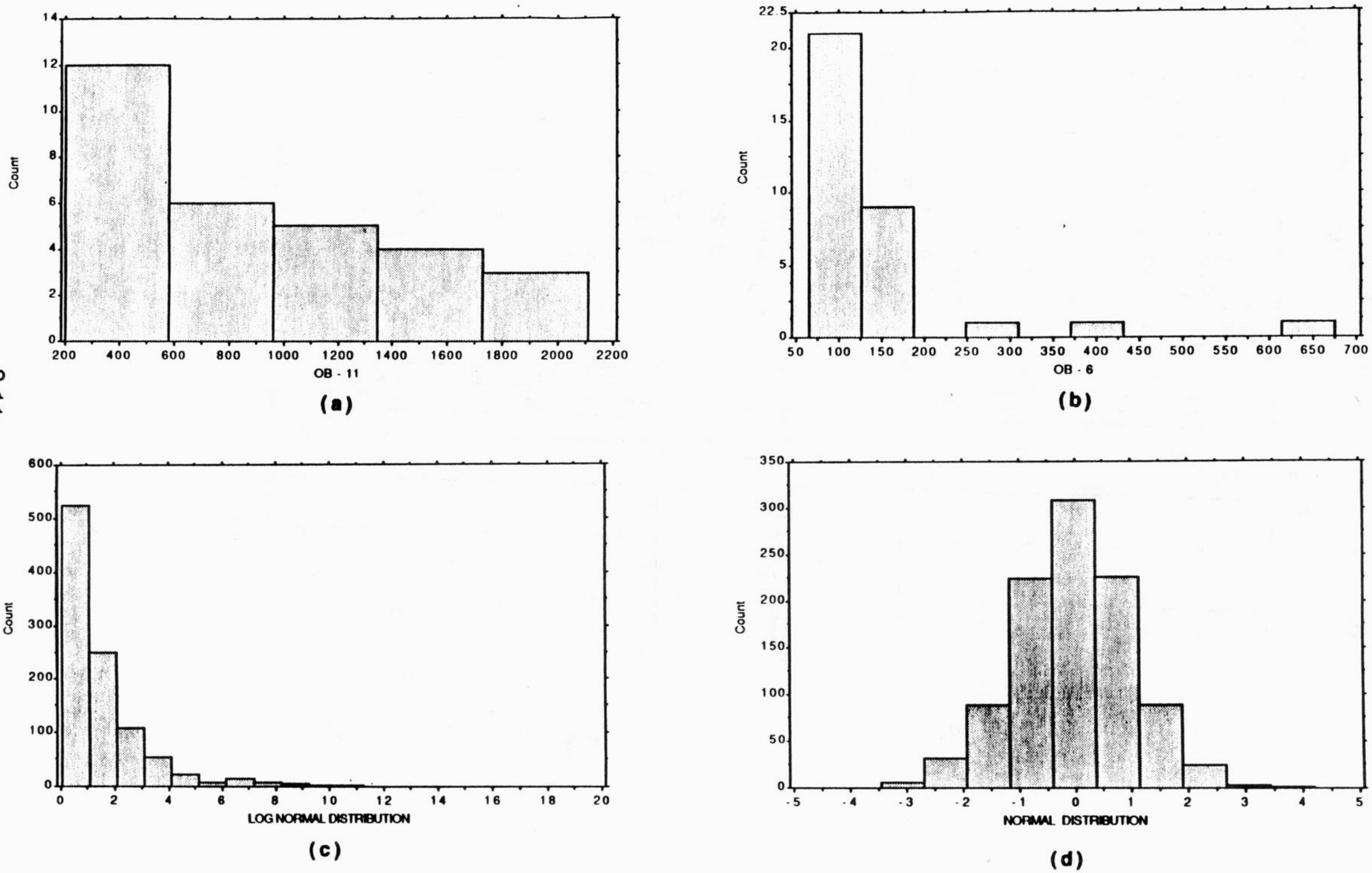


Figure 2-6 **Histograms of specific conductivity (μmhos/cm) for Wells OB-11 and OB-6 (a & b). Typical histograms for (c) the log normal distribution and (d) the normal distribution.**

FIGURE 2-7 (a) NORMAL PROBABILITY PLOT: RAW DATA

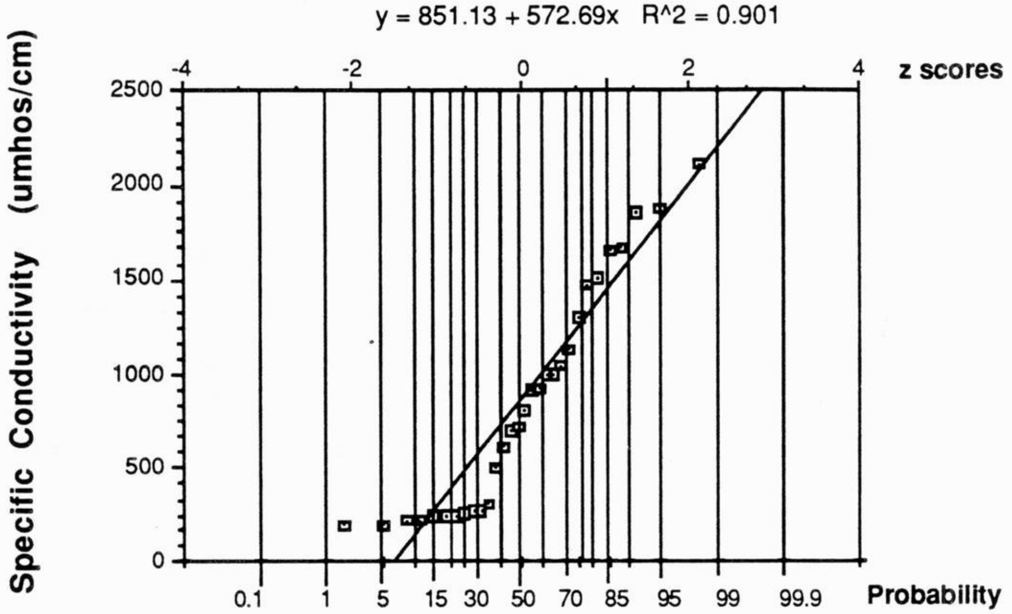
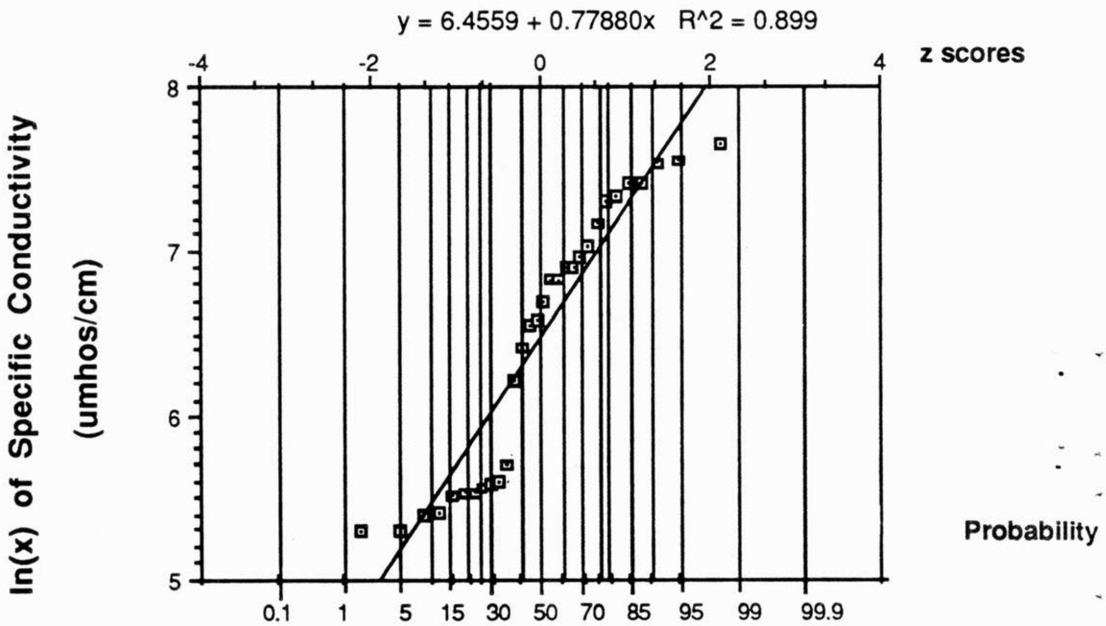


FIGURE 2-7 (b) NORMAL PROBABILITY PLOT : LOG TRANSFORMED DATA



In summary, data from the City of Merrill landfill site are presented in five ways:

(1) time versus concentration plots;

(2) box and whisker plots;

(3) one standard deviation error bar charts;

(4) histograms; and

(5) normal probability plots.

The first two graphical tools clearly illustrate qualitatively the relative water quality between wells. The error bar charts may be valuable when working with parametric statistics, particularly when deciding on PAL levels. Histograms may be used to view the probability distribution of the data. When evaluating the assumption of normality, normal probability plots are commonly prepared to observe deviations from normality.

The City of Merrill data illustrate several important points.

- Data outliers tend to inflate the mean and standard deviation of the data;
- The median and interquartile range are good estimates of the central tendency and variation of data sets, particularly when outliers are present.
- Large data variability (IQR) is usually associated with high medians, i.e. impacted wells . Natural temporal variability is much lower than the variability observed when contamination is present.
- The histograms and normal probability plots show that groundwater data may not be normally distributed. In this situation the median and IQR may be better estimates of the central tendency and variability of the data.

In the following section, the issue of non-normality is addressed in detail. Also cyclic trends in data (usually seen as seasonality) and serial correlation are investigated.

2.2 Statistical Nature of Groundwater Quality Data

While graphical tools are invaluable in conceptually understanding water quality, the application of statistics to the decision making process requires that a more quantitative determination be made of the data's structure. In order to choose a statistical test to evaluate groundwater contamination, two main factors must be considered:

- the experimental design of the test, and
- the validity of fundamental assumptions implicit in the statistical model of the test.

In this section the validity of common statistical assumptions will be explored through an evaluation of the characteristics of groundwater quality data. The issue of experimental design concerns whether a statistical test is analytically addressing the right environmental question. This is addressed in Chapter 3.

Statistical hypothesis tests are based on a model of the null hypothesis: in our case a model of background water quality. For many tests we define our model as a probability distribution and then test whether a parameter of the distribution has changed. For example the null hypothesis (see Section 1.1) for a two-sample Student's t-test is:

H_0: $\mu_{x_1} = \mu_{x_2}$, where we model background water quality by the normal distribution with parameters μ_{x_1}, the mean, and σ_{x_1}, the standard deviation (by not including σ in the null hypothesis we assume $\sigma_{x_1} = \sigma_{x_2}$). We then test whether or not the mean of downgradient water quality, μ_{x_2}, is significantly different from our upgradient model. Assumptions implicit in this model include:

- NORMALITY: The data are representative samples from a normally distributed population.
- STATIONARITY: the parameters of the probability model are not changing in time.
- INDEPENDENCE: The data are a random sample, i.e. each data point is independent of the others.

The first assumption is common to parametric statistical tests such as the t-test or analysis of variance. The second two assumptions are true for most statistical tests, including distribution-free or non-parametric tests.

The normality assumption is apt to be violated in water quality data when the data distribution is skewed. As discussed in the previous section, a histogram of the data may not resemble the normal distribution and a normal probability plot may not be a straight line. The normality assumption is tested using data from the 20 landfill sites in Section 2.2.3.

The assumptions of stationarity and independence are related to the variability found in data. In terms of statistical models, there are basically two types of variability: deterministic and nondeterministic. Fluctuations which we can explain and account for are deterministic. An example is seasonal fluctuations in mean water quality. Variability about a constant central value (i.e. the mean) which we cannot explain or explicitly account for is non deterministic. An example in this case is error introduced by laboratory analysis. It is this non-deterministic variability (sometimes misleadingly called "random noise") that we are attempting to model probabilistically.[3]

The stationarity assumption is most likely to be violated by the presence of cyclic trends, particularly seasonal fluctuations. If water quality changes by season the mean of the distribution (assumed under the null hypothesis) is not constant. This type of variability may be accounted for deterministically; i.e. we could account for this shifting mean in the t-test model by revising the null hypothesis to be:

$$H_o: \quad \mu_{x_1} + \theta_i = \mu_{x_2} + \theta_i \quad \text{for } i = 1,2,3,4$$

where μ_{x_1} is the annual mean of background water quality and θ_i is the deviation of the mean for season i from the annual mean. One approach to deal with nonstationarity would be to subtract the mean of each season from the associated data points. Thus we would reduce our seasonal model back to the original model. When there are seasonal shifts in groundwater quality data and we do not account for them, we violate the assumption of stationarity (Section 2.2.2).

[3]The idea here is that we do not know the true cause of this non-deterministic variability. It may be truly random noise or it may have some quantifiable physical explanation that we do not know.

The independence assumption is apt to be violated by the presence of serial correlation. Serial correlation is found in data which are collected too frequently to be independent of each other. When looking at time versus concentration plots, serial correlation may be a factor when high values follow high values or low follow low. In a groundwater context, serial correlation may be observed when groundwater flow is very slow, but sampling is frequent. Figure 2-8 illustrates this concept using hypothetical data with no seasonal trends (i.e. only nondeterministic variability). Serial correlation would be observed in the data set represented as circles. If sampling were done on a less frequent schedule (as shown by the squares) the assumption of independence would be valid.

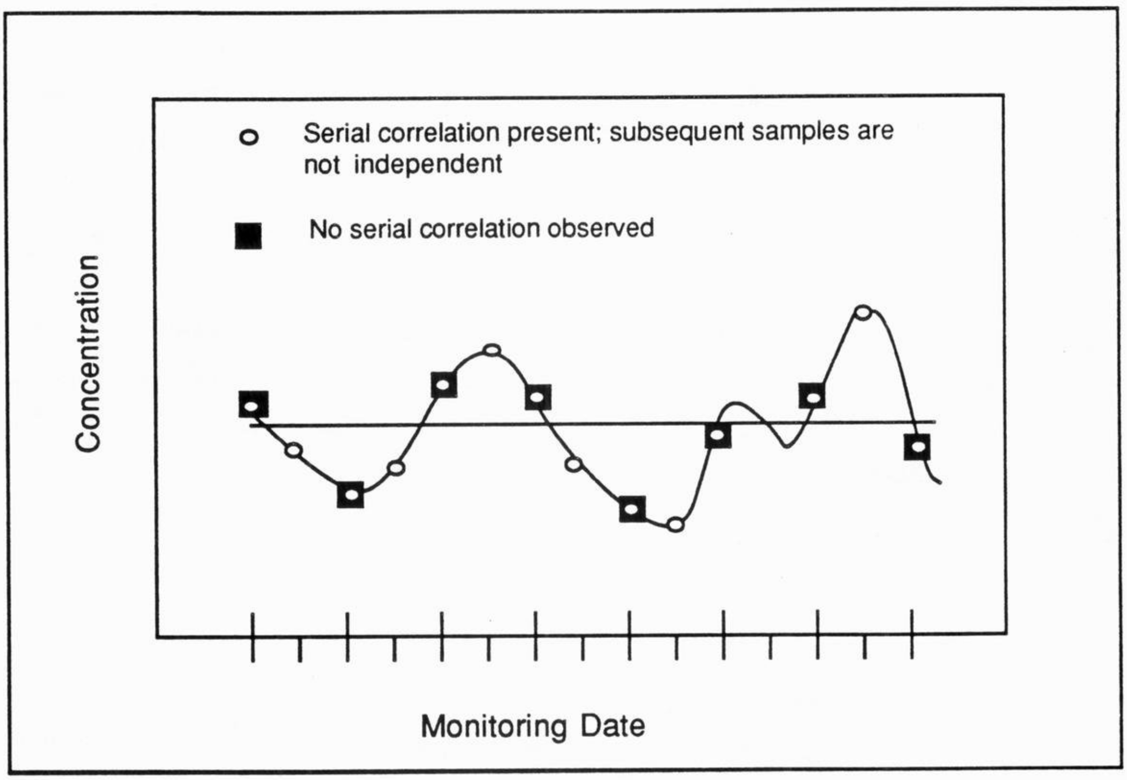


FIGURE 2-8 The concept of serial correlation in water quality data

Clearly, the presence of true trend can be confused with the presence of serial correlation. This problem is particularly acute at wastewater treatment facilities where effluent parameters are measured daily (Berthouex and Hunter, 1983).

Groundwater quality data may violate these assumptions as will be shown in the following sections. Tests of stationarity, independence, and normality are discussed in the remainder of this chapter. The assumption of normality is explored in detail since the debate between the use of parametric and nonparametric procedures hinges on this assumption.

2.2.1 The Assumption of Stationarity

Stationarity may be tested by addressing the question: do the data exhibit seasonal concentration patterns?

To evaluate the presence of seasonality the data are either (1) divided into four groups and a statistical test of "location" is applied, or (2) a time series test of periodicity is applied. Recommended tests are of both types and include parametric and nonparametric methods (Montgomery *et al*; 1987; Doctor *et al*, 1985; Harris *et al*, 1987; Montgomery and Reckhow, 1984). The most commonly used procedure is the Kruskal-Wallis test (KW). Also the Lag 4 autocorrelation function (ACF) and the one way analysis of variance (ANOVA) were used in the cited studies.

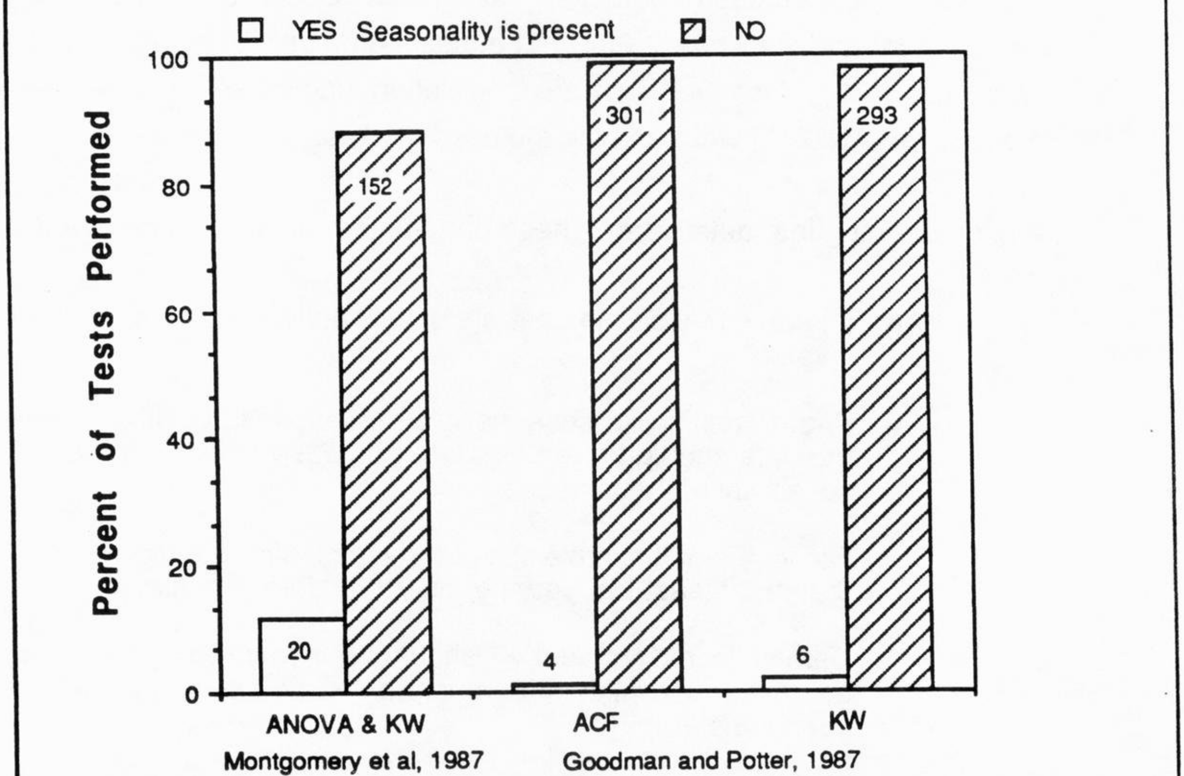
While not going into detail about these procedures, several points are important.

- If in any season the data are significantly skewed, parametric tests may be invalid.
- Prior to testing for seasonality using any test, positive or negative trends in the data should be removed (to decrease the variability and increase the sensitivity of the test; see Section 3.1.).
- For very small sample size, true seasonality is difficult to detect. As a rule of thumb, at least five years of quarterly data is minimal
- Data collected "quarterly" should be measured in the same four months each year. Monthly data must be grouped seasonally with local climatic conditions in mind.

The overall results for seasonality from two studies are presented in Figure 2-9. <u>The salient conclusion here is that groundwater quality data are not usually affected by seasonality.</u> Montgomery et al (1987) found that positive seasonality was associated with shallow unconfined aquifers with peak season usually in the summer or fall.

The importance of these findings is that it is unlikely that seasonality is a dominant contribution to the temporal variability observed in groundwater data. Many investigators have found significant seasonal influences in surface water quality and have proposed statistical methods which account for seasonal variation (Montgomery and Reckhow, 1984; Hirsch and Slack, 1984; Hirsch *et al*, 1982; vanBelle and Hughes, 1984). The low observance of significant seasonality in groundwater data suggests that these methods are rarely appropriate for Wisconsin data.

FIGURE 2-9 Overall Results for Tests of Seasonality



ANOVA = One Way Unbalanced Analysis of Variance

ACF = Autocorrelation Function

KW = Kruskal-Wallis Test

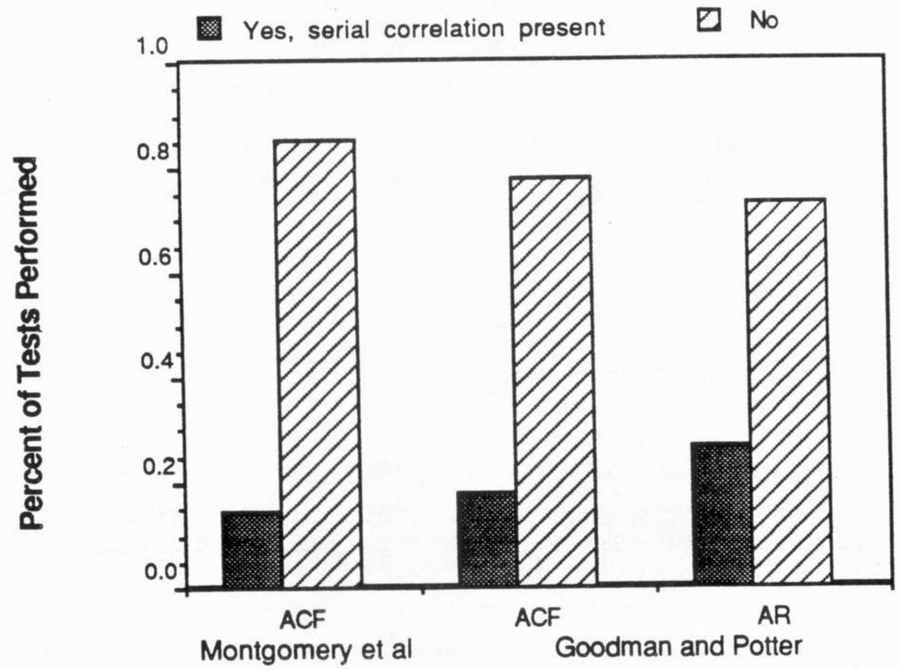
Lastly, if seasonality is believed to be a dominant source of variability, we recommend that box plots be made to illustrate the seasonal differences. Secondly, we recommend that both the Kruskal-Wallis test and the One-Way ANOVA tests be performed to confirm seasonality statistically. These are the simplest applicable procedures. The choice of both these tests is based on their simplicity and ready availability on most computer software programs, as well as on the statistical design of the test. These tests are compared and described in more detail in Chapter 3. Also, statistical procedures which take into account seasonality for evaluating groundwater contamination are discussed in Chapter 3.

2.2.2 The Assumption of Independence

Independence may be tested by addressing the question: do the data show serial correlation? From a groundwater sampling perspective, serial correlation is most likely to occur when groundwater flow is very slow; thus, concentration measurements are collected too frequently to be independent of each other. Independence can often be achieved by increasing the time between observations. Several tests have been used to evaluate the presence of serial correlation in groundwater quality data. Montgomery et al (1987) chose the Lag 1 autocorrelation function (ACF). Goodman and Potter (1987) also used this method as well as the nonparametric autorun test (AR). The application of the ACF test to groundwater quality data is described in detail by Harris *et al*, 1987. The AR test is applied to hydrologic data by Sen, 1979. Most advanced statistics texts and mainframe computer packages include these tests.

The test results for serial correlation are presented in Figure 2-10. <u>These results indicate that serial correlation may exist in groundwater quality data even though sampling is usually at three month intervals.</u> Research is needed to explore the relationship between aquifer characteristic (hydraulic conductivity, flow rate, etc...) to the statistical independence of water quality sample concentrations. Since the great majority of data sets considered did not exhibit serial correlation, We feel that <u>for everyday purposes it can be assumed that successive quarterly measurements are independent.</u>

Figure 2-10 Overall Results for Serial Correlation Tests



ACF = Autocorrelation Function AR = Autorun Test

2.2.3 The Assumption of Normality

The normal distribution is the single most important and widely used probability model in applied statistics. This is because

- many real systems fluctuate "normally" about a central mean; i.e. measurement error of a random variable is symmetric about a "true" mean and has a greater probability of being small (close to the mean) than large (in the tail of the distribution)[4]; and
- many of the parametric statistical tests are insensitive to the assumption of normality, i.e. if the data are not distributed normal, it may not matter a great deal.

[4]This fact is supported by the Central Limit Theorem which, in simple terms, states that the distribution of a sum of many small "errors" will be distributed normal, given that no single source of error dominates the rest (Box, Hunter, and Hunter, 1978).

This latter point is explored in detail in Chapter 3. Here the basic question to be answered is "Are groundwater quality data distributed normally?"

We are questioning this in the context of EPA's regulations which do not require tests for normality or other distributional assumptions unless 1) a data transformation is made, or 2) nonparametric statistical tests are applied. Many statisticians recommend data transformations to "normalize" skewed data for parametric tests. EPA regulations advocate transformations if necessary. Many environmental systems are modelled using the lognormal distribution because (1) it has a lower bound of zero, and (2) is positively skewed, allowing high values to be included (Benjamin and Cornell, 1970)[5]. Unless there is a physical justification to delete high values they must be considered as a part of the dataset.

In this section, the assumption of normality is evaluated for eight parameters monitored at 20 Wisconsin waste disposal sites. First statistical tests for normality are discussed. Then, results of the tests of normality are presented for both raw and log-transformed data. The objective of the normality analysis was to gain insight into whether or the not normal or lognormal distributions are appropriate probability models for these eight parameters. In addition EPA policy and recommended methods are evaluated.

<u>TESTS FOR NORMALITY</u>.

To test the hypothesis of normality many statistical "goodness-of-fit" tests may be used. These tests mathematically compare the shape of the normal distribution to the data set of interest. These tests should only be applied to independent, stationary data sets. As shown in the previous sections groundwater quality data in Wisconsin usually meet these criteria.

Shapiro et al (1968) did a comparative (Monte Carlo) study of nine tests for normality, evaluating the sensitivity of the tests to small sample size. At small sample sizes it is hard to reject normality even if the data are not normal. A sensitive test is one which can detect non-normality even at small sample size. Shapiro et al (1968) found that

[5]The model for the lognormal distribution is $Y=\ln(x)$, where x = original concentration. x is said to be distributed log normal. "ln" is the abbreviation for the natural log of x.

- The W-statistic (Shapiro and Wilk, 1965) was preferred.
- The Kolmogorov-Smirnov and chi square test -- the most often used distribution tests -- were relatively insensitive.
- And, a combination of the third sample moment (skewness) and fourth moment (kurtosis) provides a sensitive judgement, but even their combined performance is less than his W-statistic.

The W-statistic however is not used widely because it is not readily available on computer software.

In the groundwater quality literature, Montgomery et al (1987) tested the normality of groundwater quality data using graphical methods, the chi square test and the skewness test. Harris et al (1987) recommend the skewness test for general use with groundwater quality data. EPA's draft guidance manual for statistical analysis at RCRA facilities recommends three statistical procedures to check normality:

1) the chi square test,
2) the coefficient of variation method, and,
3) normal probability plots.

As mentioned above the chi square test does not perform well at small sample size (less than 20). Furthermore, the method is very conservative for continuous distributions such as the normal.[6] For these reasons, we do not recommend the chi square test for general use with groundwater quality data from waste disposal facilities.

The second EPA method relies on the sample coefficient of variation, ν, which is equal to the sample standard deviation divided by the sample mean. This method was previously required by EPA and is expected to be used widely (EPA, 1988). The rule is: if ν is greater than 1 do not assume normality. The idea is if $\nu = s/\bar{x}$ is greater than 1, then the normal probability model will predict negative concentration values with an (unacceptable) high probability. Water quality concentrations are inherently non-

[6]Conservative implies that the test will not reject the normal when in fact it should; i.e. the test will have a high Type II error rate.

negative. An evaluation of this method is made below by comparing results to the results of the skewness test.

The normal probability plot is a qualitative method, and thus cannot provide statistical inference.[7] We do not recommend any of EPA's suggested tests for normality, although normal probability plots are useful for illustrating deviations from normality.

The results presented below are based on the skewness test. We recommend the skewness test because:

- the coefficient of skewness is easy to calculate and is included in all statistical software packages;
- the test is simple, requiring only a comparison of the skewness coefficient to tabulated values (see Appendix A);
- the critical levels (table values) have been generated for small sample sizes (Harris *et al*, 1987); and
- the test has been found to be robust at small sample size by Shapiro and Wilk, 1965.

The procedure for applying the skewness test to a data set is briefly described in Appendix A.

SKEWNESS ANALYSIS.

In this study sample data for eight parameters from 161 groundwater wells located at 20 landfill sites are tested for normality using the skewness test and, for comparison purposes, the coefficient of variation method. Results are used to evaluate the general use of parametric statistical tests. Also, the normal and lognormal distribution are each evaluated as being in general an appropriate probability distribution for each parameter. Only wells with data representative of background water quality are included (161/274 wells from the 20 sites). Only background datasets were tested because samples obtained in a contaminated situation are most likely not drawn from a single population (that is, since leachate plumes are not well-mixed and evenly distributed in space) See Section 4.1 for a discussion of how background water quality wells were distinquished

[7]There is a test based on the correlation coefficient of the probability plot. See Vogel, 1986 (Water Resources Research. Vol. 22 No. 4: 587)

from impacted wells. Each data set consists of the time series of concentrations measured at a well for one parameter. Only data sets with zero samples reported at or below the analytic detection limit and only data sets with nine or greater samples are included in the analysis. A total of 699 concentation time series from the 161 wells met these criteria. The skewness test is applied to both the raw data and the log-transformed data. A two sided test for positive or negative skew is applied at $\alpha = 0.10$. The hypotheses are:

$H_0: |\gamma| < \gamma_{n, \alpha = .05}$ Sample may be from a normal distribution.

$H_1: |\gamma| \geq \gamma_{n, \alpha = .05}$ The normal distribution is rejected.

where γ is the skewness coefficient as defined on Table 2-2.[8] Values for $\alpha_{0.05}$ are tabulated by sample size in Appendix A. In general, if skew is greater than 1.0 the data are found to be non-normal regardless of sample size.

Overall results are compared to two previous studies in Figure 2-11. Note that data analyzed by Goodman and Potter (1987) is a subset of the data considered in this study; however, Goodman and Potter did not eliminate datasets with observations at or below the limit of detection. In this study, 47 percent of the datasets were found to be non-normal. The lognormal distribution failed to fit 43 percent of the datasets. For those 238 data sets which rejected the normal distribution the lognormal distribution was *not rejected* for 105. Thus either the normal or lognormal distribution was found to "fit" 68 percent of the data sets. However, the results show that groundwater quality data frequently violate the assumption of normality.

The results from individual parameters are shown in Figure 2-12 and summarized in Table 2-3. This summary shows:

- pH is least apt to be skewed. Since pH is already on a logarithmic scale this result is not surprising.
- Conductivity and alkalinity data were significantly non-normal less than 45 percent of the time.
- COD, iron, and chloride data are most frequently skewed; this may be

[8]Note that these hypotheses imply an overall $\alpha = 0.10$ since the absolute value of γ is being considered. That is, we are not assuming apriori that skew is positive or negative.

FIGURE 2-11 Overall Skew Results

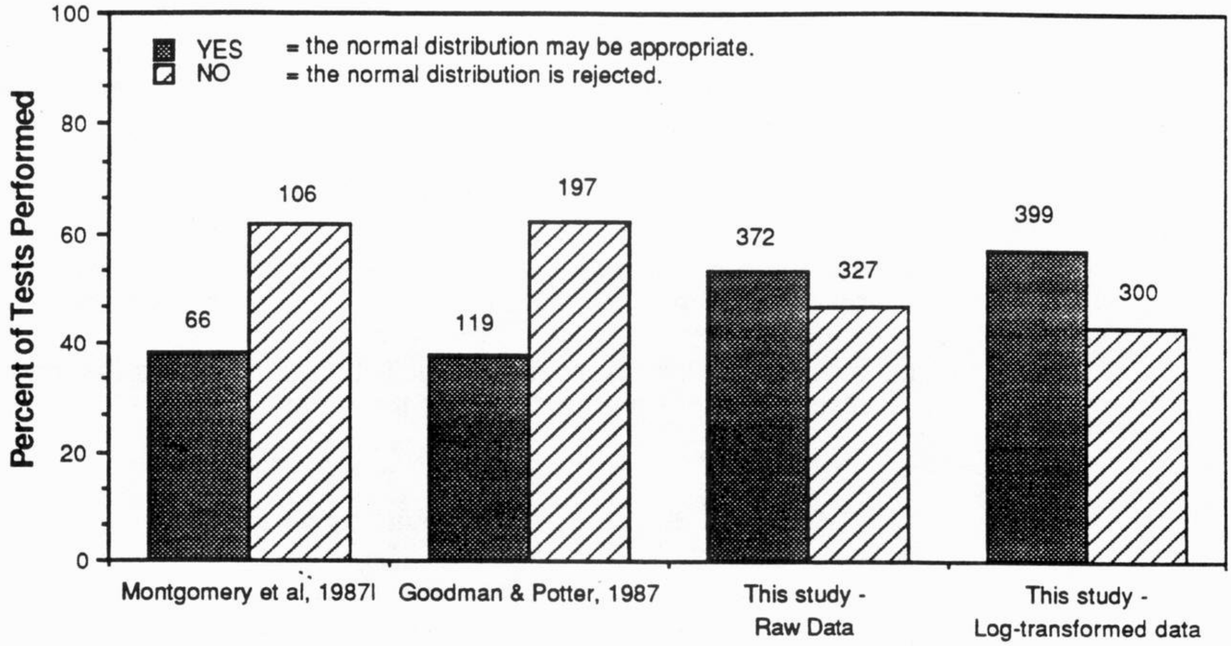
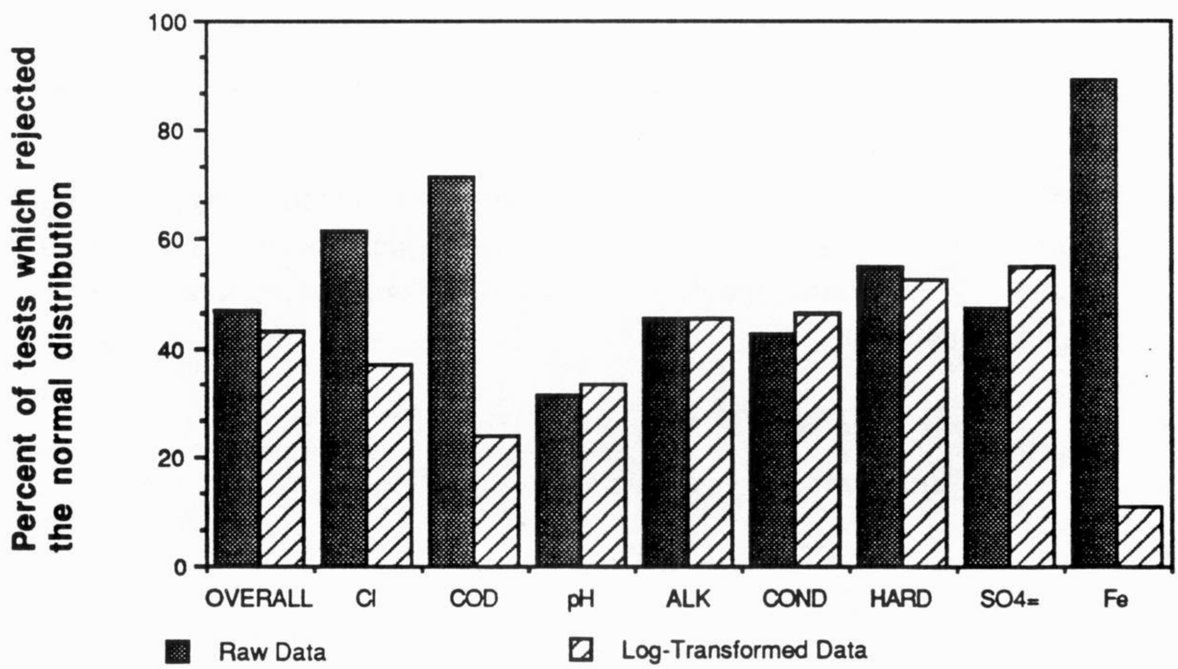


FIGURE 2-12 Parameter Results of Skewness Test for Normality for Raw and Log-transformed Data



because these parameters are often at or near the analytic detection limit in background water quality.

Comparing the raw and log-transformed data, Figure 2-12 illustrates that

- the lognormal distribution reduces the frequency of positive skew for most parameters, and particularly for iron, chloride and COD; and,
- for pH, alkalinity, specific conductivity and hardness the lognormal distribution performed similar to the normal.

TABLE 2-3 Skewness test results by parameter

	Number of Non-normal Test Results	
PARAMETER	Raw Data	Log-Transformed Data
Chloride	47/76	28/76
COD	15/21	5/21
pH	47/150	50/150
Alkalinity	54/119	54/119
Conductivity	65/153	71/153
Hardness	72/131	69/131
Sulfate	19/40	22/40
Iron	8/9	1/9

Figure 2-12 implies that the lognormal distribution is more appropriate in general than the normal. To explore this idea further, the results were divided into four categories depending on whether the distributions were/were not rejected at the 5 percent significance level.

1) both rejected

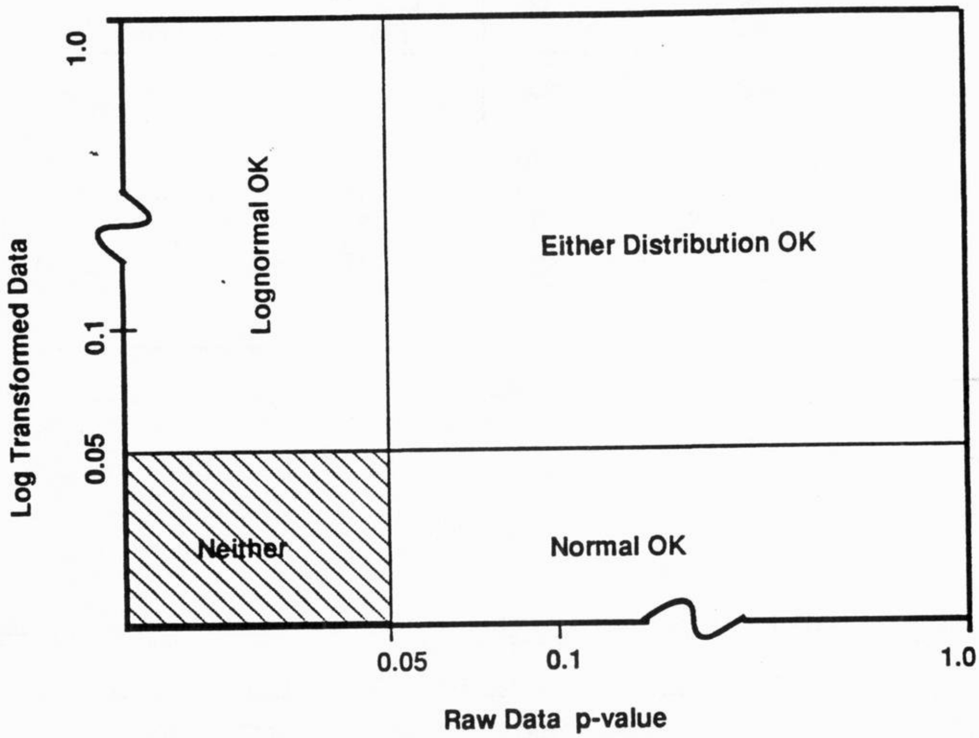
2) only lognormal rejected

3) only normal rejected

4) neither rejected.

If only the lognormal distribution is rejected, then the statement may be made that the normal distribution appears to better represent the data. Figure 2-13 illustrates these four possible outcomes. If it had been possible, the exact p-values for each test would have been plotted on a figure similar to Figure 2-13. However, available tables are limited to

FIGURE 2-13 Possible Outcomes for Skewness Test of Normality



specific significance levels. The skewness results are divided into the four groups above for each of the eight parameters. These results are summarized in Table 2-4.

Table 2-4 shows that

1) more than 20 percent of the datasets were found to reject both the normal and lognormal distribution (except for iron).

2) For pH, alklainity, conductivity and hardness, both distributions were found to fit more than 1/3 of the data sets. These parameters are found to have less natural variation in groundwater (relative to the mean) than the other parameters. Since the lognormal distribution is similar to the normal when variance is low, these results are not surprising.

3) As also shown by Figure 2-12, datasets for chloride, COD and iron rejected the lognormal distribution less frequently than the normal distribution.

TABLE 2-4 Comparison of Skewness Test Results for the Normal and Lognormal Distribution

PARAMETER	Total	Log-normal Preferred	Normal Preferred	Neither	Either
Chloride	76	27	8	20	21
COD	26	10	0	5	6
pH	150	3	6	44	97
Alkalinity	119	12	12	42	53
Conductivity	153	18	23	47	65
Hardness	131	18	14	54	45
Sulfate	40	9	12	10	9
Iron	9	8	1	0	0

Unfortunately these results do not imply that one distribution is preferred over the other for any one parameter. In fact, the salient conclusion here is that the assumption of normality should always be tested for before parametric statistics are applied to groundwater quality data.

EPA (1988) advocates that it is protective of the environment to adopt the appoach of not requiring testing of assumptions on a wide scale because "only extreme violations of assumptions will result in an incorrect outcome of a statistical test." The results presented thus far do not yield any insight into the "degree of violation." To investigate this issue further, the distribution of the skewness coefficient was inspected for each parameter. The skewness coefficient for large samples drawn from a normal population is distributed normally with $\mu = 0$ and $\sigma = 6/N$, where N is sample size.[9] Thus one approach to investigate the "degree of violation" of the normality assumption would be to standardize

[9]The distribution of the skewness coefficient is independent of the mean (μ) and standard deviation (σ) of the population from which the samples are drawn. Thus, coefficients calculated at different wells in different geologic strata could be from a single normal population (that is, after the coefficients are standardized: see text).

the skewness coefficients by dividing by the expected value of the standard deviation, $\sqrt{6/N}$. The resulting dataset for each parameter should be distributed standard normal. The same approach is applicable for the log-transformed data to evaluate departure from lognormality. Figure 2-14 (a) through (h) are plots of the standardized skewness coefficient for the raw data versus the log transformed data. All points greater than 1.64 units from the origin will be found to reject the normal distribution at the 5 percent significance level (for $F(x) = 0.95$, $z=1.643$)[10]. These plots may only be roughly interpreted since the distribution of the skewness coefficient at small sample size is not exact. Figure 2-14 shows for the normal distribution that many parameters include points which may be considered extreme violations as evidenced particularly for conductivity, total hardness, chloride and COD. For all parameters the lognormal distribution appears more symmetric about the origin and has fewer extreme violations than the normal distribution. Thus, while we cannot state that the lognormal distribution is always the "best" choice for these parameters, it appears to be a better first choice than the normal distribuiton.[11]

In summary, based on the skewness test we have found that 53 percent of the raw data sets are approximately "normal." An additional 15 percent are found to be approximately normal after log transformation. These results are important because they show that parametric statistical tests may not be valid in many cases. The results for individual parameters do not imply that either distribution is "best" for characterizing clean water quality. However, inspection of the distribution of the skewness coefficient implies that the lognormal distribution may be a better first choice than the normal for most parameters. The implication of these results is that parametric tests must be used with caution.

Presented below is a comparison of the skewness test to EPA's recommended coefficient of variation method.

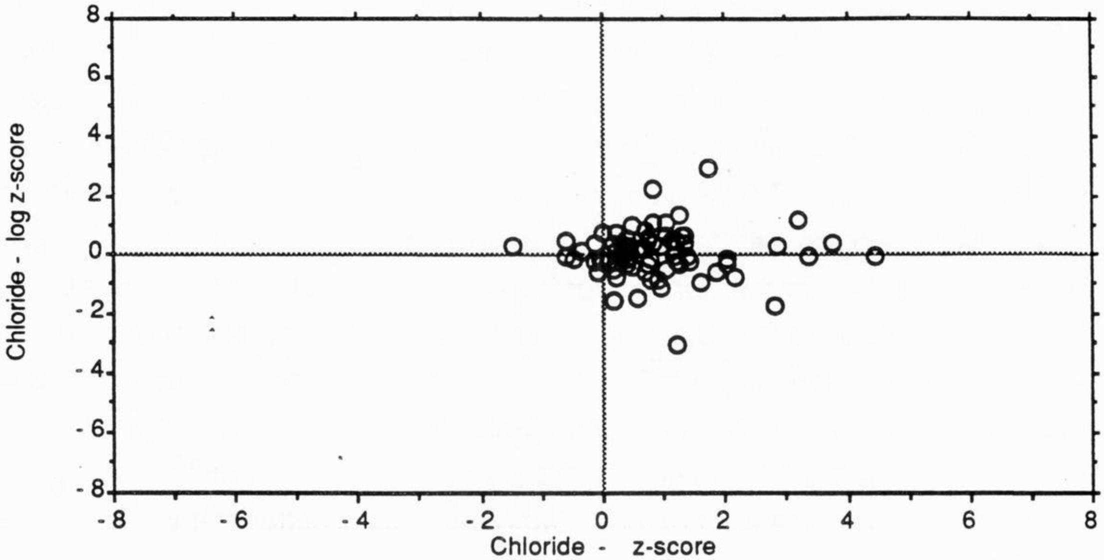
---

[10] $F(x)$ is the cumulative distribution function of the standard normal distribution; $z = 1.643$ implies that 95 percent of the standard normal distribution is less than 1.643 standard deviations from the mean.

[11] At first glance, these results for pH are surprising since pH is already on a logarithmic scale. However, as mentioned above with respect to Figure 2-14, pH data is usually found to have low variance, and thus, the lognormal distribution closely resembles the normal.

FIGURE 2-14 Distribution of Skewness Coefficient for Raw and Log-transformed Data

(a)



(b)

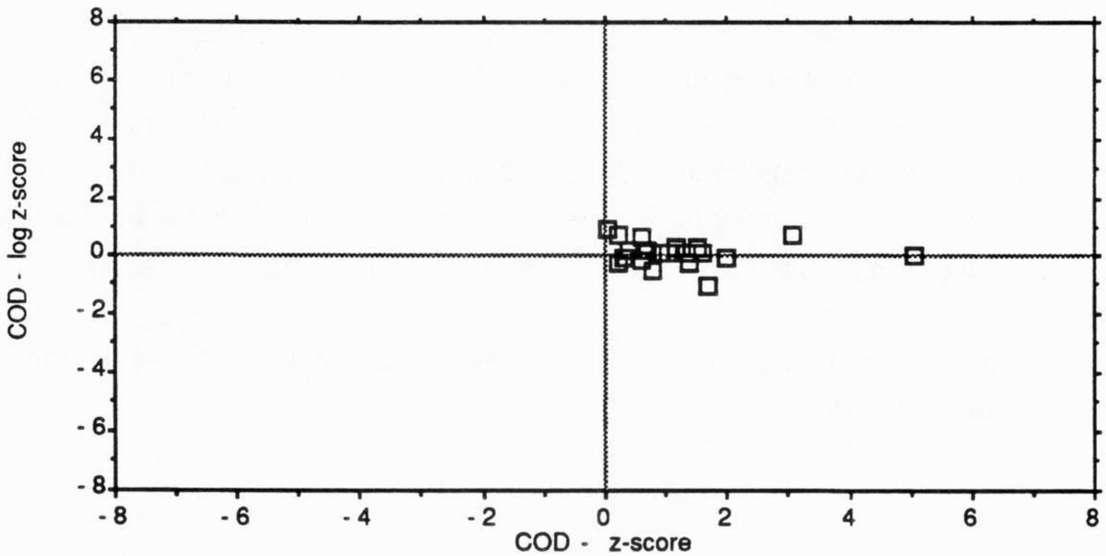
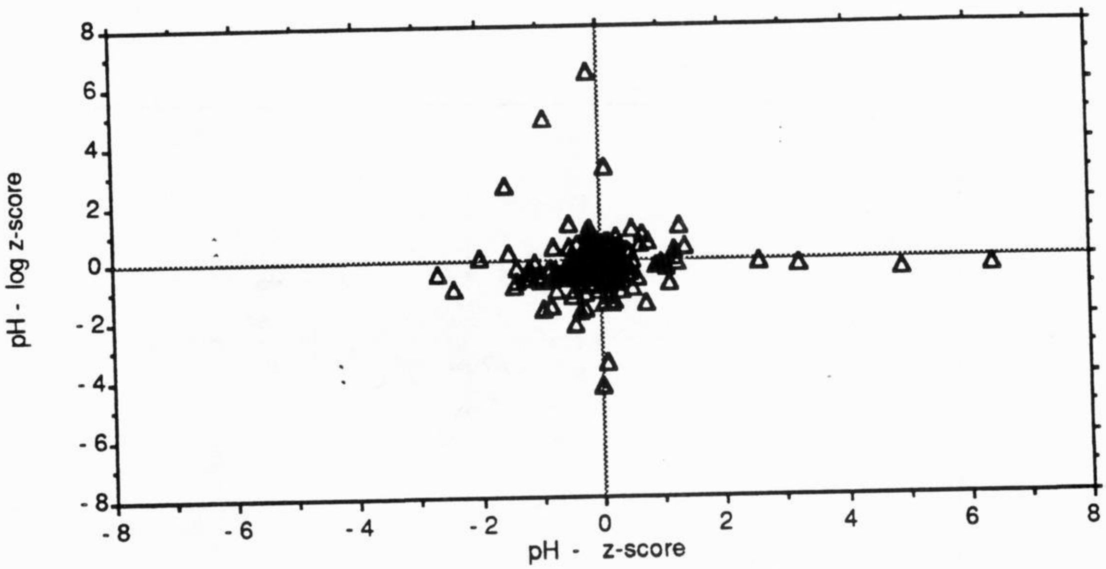


FIGURE 2-14 (continued)

(c)



(d)

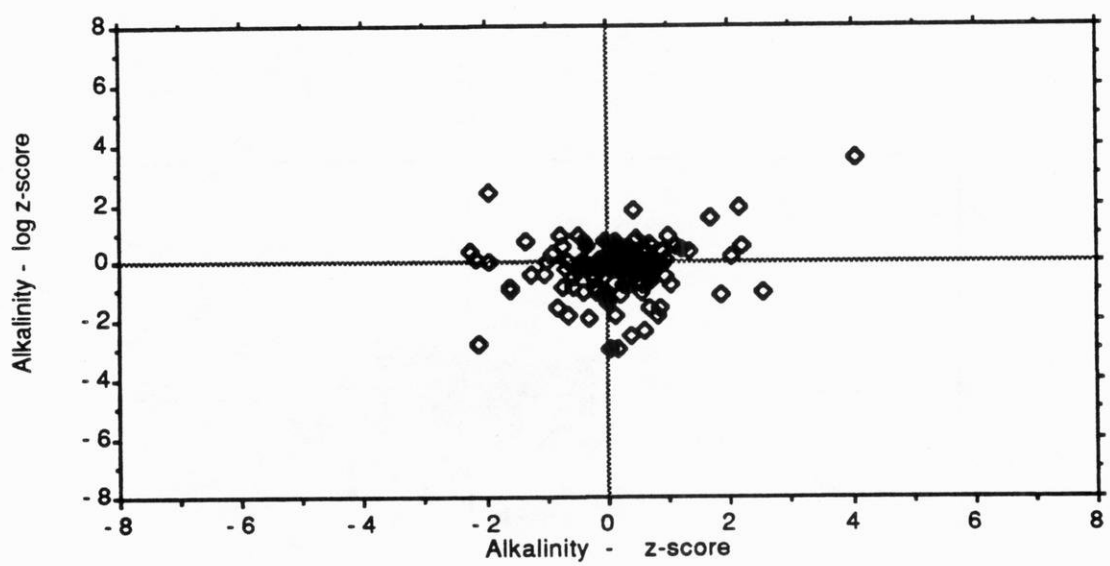
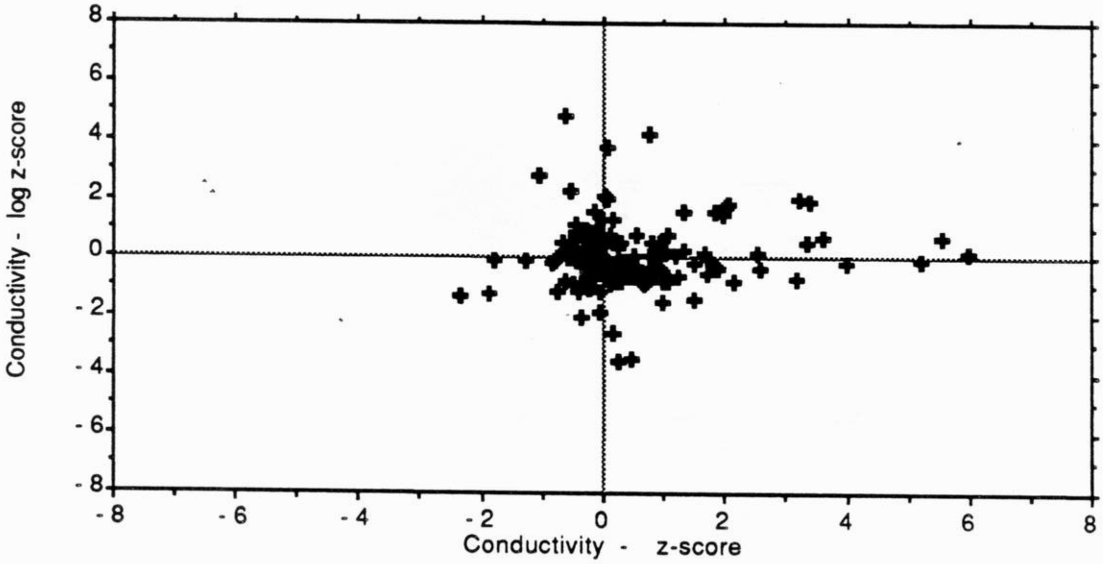


FIGURE 2-14 (Continued)

(e)



(f)

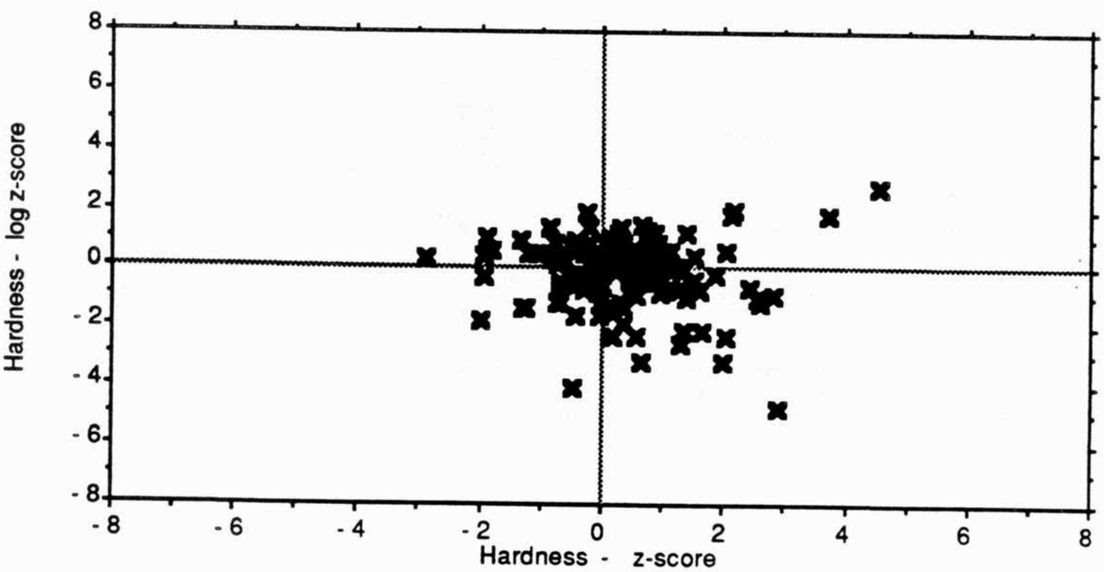
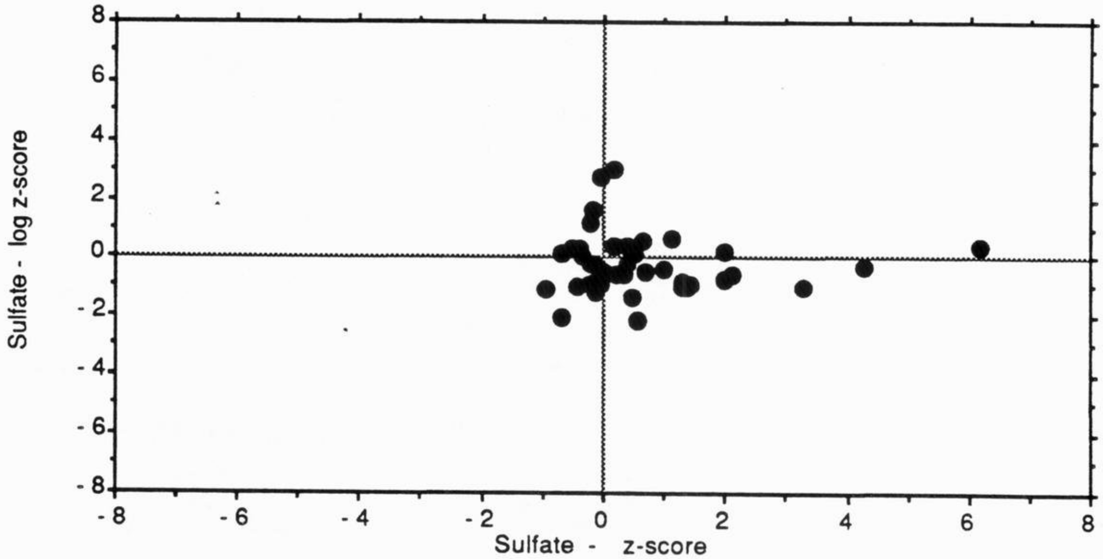
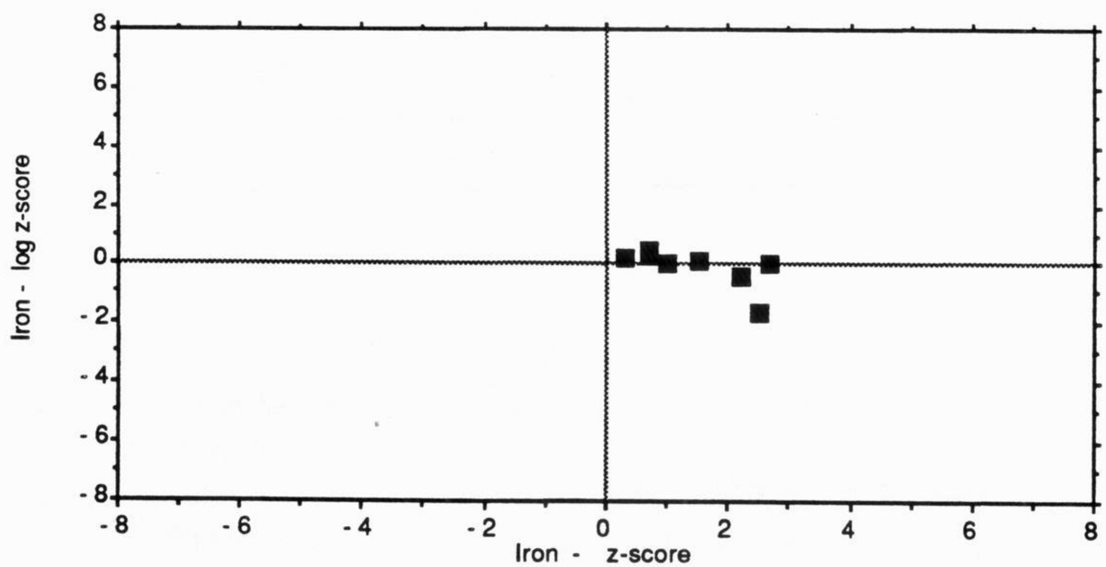


FIGURE 2-14 (Continued)

(g)



(h)



COMPARISON OF SKEWNESS TEST TO EPA METHOD

The skewness test results presented above are compared to the coefficient of variation method recommended by EPA. Of the 699 raw datasets considered in this study only 5 percent had a coefficient of variation greater than 1.0. Remember that for 47 percent of the same datasets, the assumption of normality was rejected based on the skewness test. The coefficient of variation method grossly underestimates the number of non-normal datasets Results for specific parameters are listed in Table 2-5.

TABLE 2-5 Comparison of Skewness Test to Coefficient of Variation Method

PARAMETER	NUMBER OF NON-NORMAL TESTS SKEWNESS TEST RAW DATA	COEFFICIENT OF VARIATION METHOD
Chloride	47	7
COD	15	7
pH	47	4
Alkalinity	54	0
Conductivity	65	2
Hardness	72	2
Sulfate	19	4
Iron	8	7
TOTAL	327/699	33

Based on these results we do not feel that the coefficient of variation method should ever be used as an indicator of normality. In concept it only "protects" against the prediction of negative values (for example, when setting prediction limits or tolerance intervals). The assumption of normality must be evaluated by a "goodness-of-fit" test.

2.2.4. Summary

The statistical characteristics of groundwater quality data were evaluated in order to test the validity of fundamental assumptions implicit in statistical hypothesis tests. The results show that the parametric assumption of normality is violated quite frequently while the assumptions of independence and stationarity are not. Seasonality and serial correlation were evaluated to respectively test the assumptions of stationarity and independence. Both assumptions were found to be generally valid. On the other hand, the assumption of normality was found to be violated 47 percent of the time. Even with a log transformation the normal distribution was rejected 43 percent of the time. Overall, 68 percent "fit" either the normal or log normal distribution. These results show the importance of testing for normality before applying parametric statistical tests. The wide occurrence of non-normality supports the use of nonparametric statistical procedures.

A comparison of EPA's coefficient of variation method to the skewness test results showed that the coefficient of variation method is inadequate: the method estimated that only 5 percent of the datasets were non-normal compared to 47 percent for the skewness coefficient.

2.3 Regulatory Perspective

In this chapter the numerous sources of variability and error in groundwater quality data were introduced. Graphical methods were illustrated. In addition, we showed that the assumption of normality is often violated, while the assumptions of stationarity and independence are generally valid. To place this work in a regulatory context three situations are briefly considered:

- applicability to everyday activities at the DNR;
- applicability in a court of law where groundwater quality regulations are to be enforced; and
- applicability to facility owners/operators and their consultants working on a quarterly basis to meet DNR and/or EPA requirements.

Graphical displays are by far the most valuable and essential tools in all three of the above situations. A case of clear groundwater contamination by a landfill may be built solely with graphical data presentation, explanation of geology and standard exceedances. It may not be necessary to apply statistical tests.

Whenever parametric statistical procedures are to be applied, DNR staff should always consider the validity of the assumption of normality. The skewness test is recommended. This method is briefly described in Appendix A. Because seasonality and serial correlation were not found to be prevalent, we recommend that tests for these factors be applied only in special cases.

In a court of law clear graphical display of supposed contamination is essential. If a court case is being built using results of statistical tests (as well as hydrogeologic information, etc..) then the validity of the assumptions underlying the tests may be a central issue. We recommend the following methods:

NORMALITY:	Normal probability plots The skewness test
STATIONARITY (SEASONALITY)	Box plots with the Kruskal-Wallis test
INDEPENDENCE (SERIAL CORRELATION)	The Lag 1 Autocorrelation function or the Auto-Run test.

At RCRA waste sites a quarterly test for "change" in water quality is required (or proposed to be required at MSWLF's) and time versus concentration plots must be submitted. Tests for distributional assumptions are required only when 1) a data transformation is made or 2) nonparametric statistical tests are employed. Wisconsin does not require tests of distributional assumptions. Of the three methods recommended by EPA in the draft guidance manual for statistical analysis at RCRA facilities, we feel that only the normal probability plots are useful. The chi square test and coefficient of variation method are inadequate in many cases. We recommend that the skewness test be suggested for general use because it is sensitive to small sample size, and may be performed quickly and easily.

The prevalence of non-normal datasets complicates the choice of statistical test, since nonparametric methods may be more appropriate than the traditional parametric tests. Chapter 3 explores this issue. At this point, we suggest that

1) DNR require graphical summaries of site water quality. These could be prepared by solid waste site owners on a regular, perhaps annual, basis. A quick review of time versus concentration graphs may show trends in time and any abrupt changes in water quality; and,

2) DNR require that any statistical analyses (submitted by owners/operators for setting PAL's or for detecting "change" in water quality) include tests for normality to justify the use of parametric methods before applying them.

CHAPTER THREE
EVALUATION OF GROUNDWATER CONTAMINATION

3.0 OVERVIEW

Enforcement of groundwater quality regulations at waste disposal facilities requires not only a determination that contamination exists but also evidence that it is due to the facility. Exceedance of water quality standards is largely relied on as an indication of contamination. Yet even a standard exceedance must be compared to background water quality to conclude that the facility owner is responsible. Thus, comparison of downgradient water quality to "known" background water quality is an important regulatory strategy. In this chapter we address the issue of how to detect significant changes in water quality given the natural temporal and spatial variability in background water quality. Statistical tests currently recommended by EPA and other methods proposed in the water quality literature are evaluated. Examples are drawn from the 20 sites considered in this study to illustrate the use and misuse of these statistical tests. Recommendations are made for analysis of two general situations:

- existing municipal solid waste landfills (MSWLF's),
- new facilities and existing facilities with historically clean water quality,

While MSWLF's are the focus of this study, the recommendations developed are also applicable to hazardous waste disposal sites, industrial waste disposal sites, land disposal sites for wastewater, and similar situations. The recommended methods are synthesized into general procedures presented by flow charts in Chapter 5.

The chapter is organized as follows.

- General types of tests and the questions they answer are introduced in this section, followed by a general comparison of parametric and nonparametric tests.
- Evaluation of statistical tests follows in Sections 3.1, 3.2, and 3.3.
- Conclusions and recommendations are summarized in Section 3.4 and discussed from a regulatory perspective in Section 3.5.

The overall objective of this chapter is to highlight some of the theoretical and practical limitations of specific statistical tests and to recommend tests to DNR. More specific objectives are to provide DNR with:

- guidance on establishing background water quality for setting preventive action limits (PAL's);
- advice on using statistical tests to determine exceedance of a water quality standard; and
- specific tests to determine significant changes in background water quality.

3.0.1 Types of Statistical Tests

Four general categories of statistical methods are currently considered appropriate for determining compliance with groundwater quality regulations:

- tests of central tendency (location),
- tests of trend,
- prediction, tolerance, and confidence intervals, and
- control charts.

Tests of central tendency compare whether or not the mean or median of two or more datasets are significantly different. Tests of trend look for significant increases or decreases in water quality over time. Prediction and tolerance intervals are methods which set brackets for "acceptable" background water quality based on existing data. Confidence intervals are brackets for "average" background water quality. Control charts are graphical methods widely used in industrial engineering and are similar to the intervals mentioned. This study focuses on the first three categories. Control charts will not be addressed.

While it is the statistician's job to develop correct procedures for answering relevant questions, the engineer/scientist must decide on the relevant questions. Once the right questions are clearly stated, it should be easy to decide which type of test is appropriate (if any).

In order to select an appropriate test we must address the following issues:

- What is the right question?
- Which tests have the appropriate statistical model to answer the question?
- Do our data violate the implicit assumptions of the model?

After application of a test, it is essential to evaluate if the results are meaningful.

- Do plots of the data support the statistical results?
- Are statistically significant results environmentally meaningful?

From a regulatory perspective the two questions generally asked are:

- Is a groundwater quality standard exceeded?
- Has water quality significantly changed?

These questions however are not specific enough to choose a type of statistical test. Consider Figure 3-1, a site map for Wausau Paper Mills sludge landfill. Wells P-7 and P-3 are upgradient of the disposal cells. Wells P-8 and P-9 are between Cell 2 and Cell 3. Wells P-1 and P-4 are downgradient of the disposal area. Monitoring has existed at this site since late 1981; however wells P-7, 8 and 9 were not installed until late 1984. One sample is collected from each well quarterly. Specific conductivity data are plotted versus time on Figure 3-2 and as box plots on Figure 3-3. These figures illustrate that the two downgradient wells have historically higher concentrations than the other four wells. Similar results are found for many other parameters. Clearly, the disposal area is contributing to these elevated concentrations.

What are the questions of concern at this site? Are statistical tests necessary to document groundwater contamination? At existing sites with apparent contamination a possible question is

"Has this site *historically* affected groundwater quality?"

Evaluation of groundwater flow and geology, together with graphs of water quality, may clearly show contamination. To answer this question statistically a multivariate test of central tendency such as a parametric or nonparametric analysis of variance (ANOVA) may be appropriate (Section 3.1). These tests are designed to evaluate whether or not

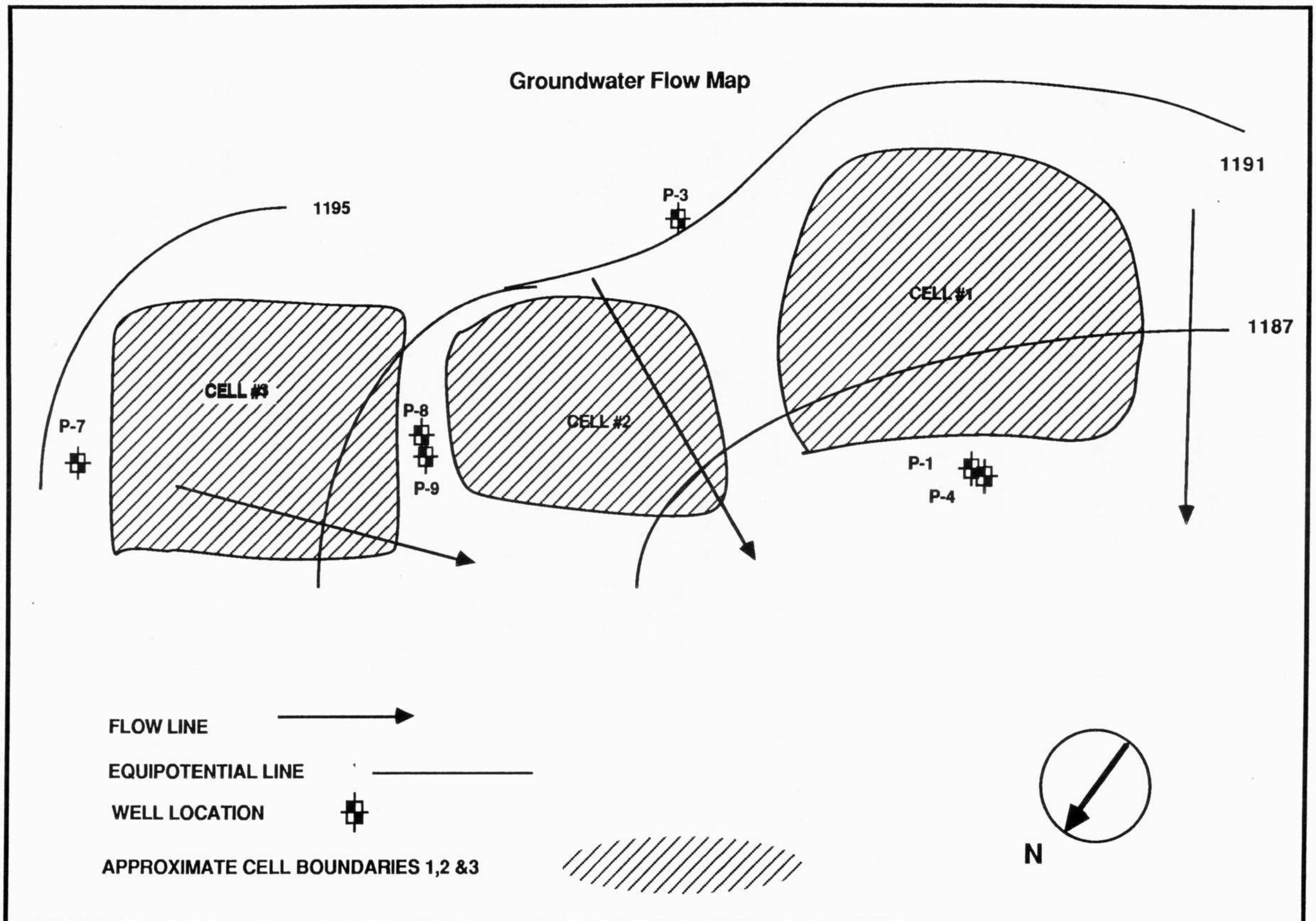


FIGURE 3-1 Site map for Wausau Paper Mills landfill

FIGURE 3-2 Specific Conductance vs. Time at Selected Monitoring Wells.

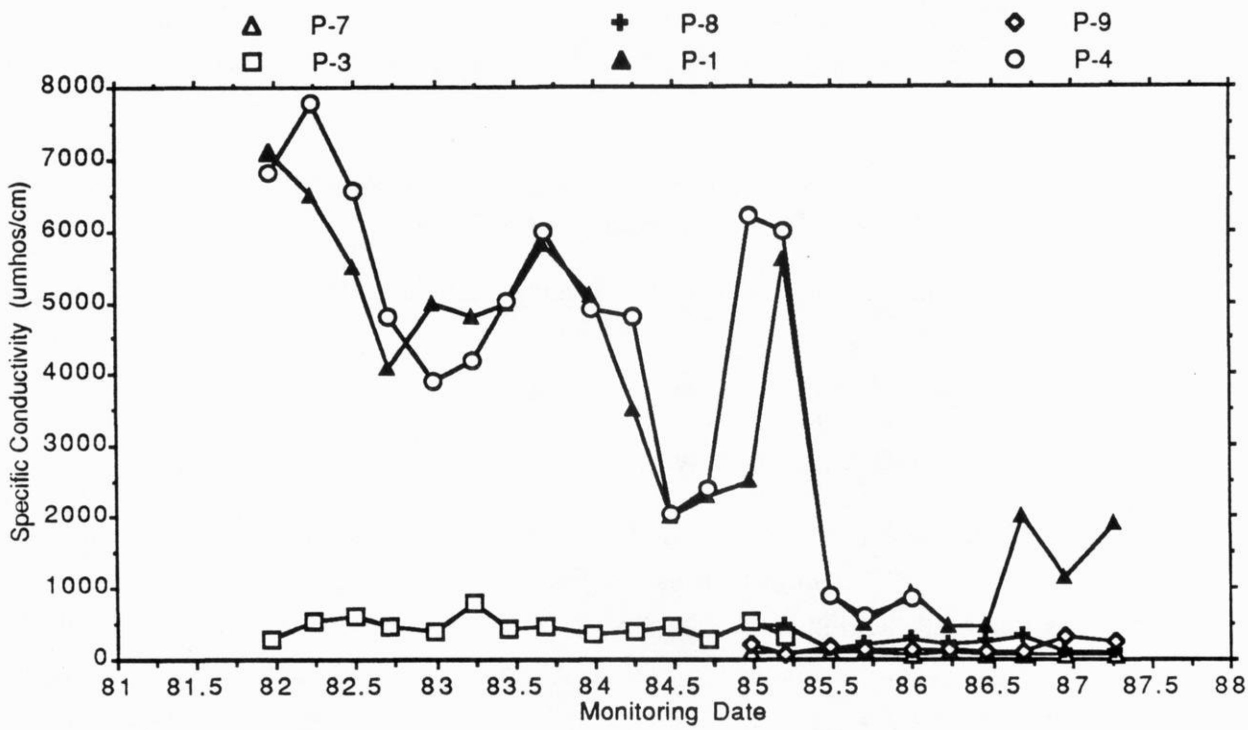
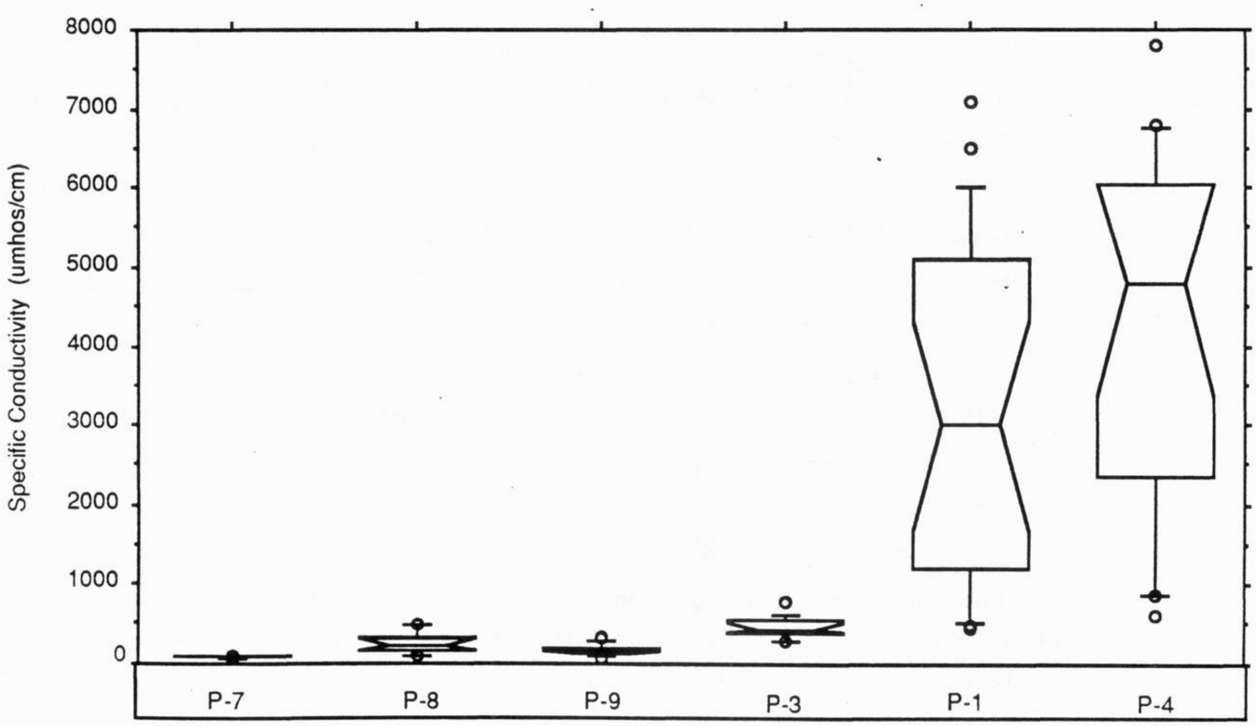


FIGURE 3-3 Boxplots for Specific Conductance at Selected Monitoring Wells



a significant difference exists between the historical mean/median of background water quality and the mean/median of each downgradient well. As shown further on, spatial variability in background may confound the results. The above question could be rephrased to

> Do concentrations observed at downgradient wells fall within limits established for average background water quality?

This question could be addressed using statistical confidence intervals (Section 3.3).

Another possible question is:

> "Has water quality improved since monitoring began?"

In this case, a test of trend over the length of record may be appropriate. Trend analysis could include parametric linear regression or a nonparametric measure of trend such as Kendall's Tau statistic (Section 3.2). Inspection of Figure 3-2 for wells P-8 and P-9 suggests that the answer to this question is "yes." Now, for discussion purposes only, assume that Cell 3 is a separate facility from Cell 1 and 2. The question of concern is now:

> "Has Cell 3 contributed to contamination of groundwater?"

The historic data from Wells P-8 and P-9 suggest that Cell 3 is not impacting groundwater. If this is so, we are in the situation where we have a currently "clean" site, and we are interested in assessing whether new data at Wells P-8 and P-9 reflect a change in water quality. A test of central tendency is no longer appropriate since historic water quality is clean. We are no longer interested in comparisons of average (mean) water quality. [1] Rather we want to compare recent data to historic background water quality. When we want to compare a single round of new measurements to background, prediction intervals, tolerance intervals and control charts may be appropriate (Section 3.3). Another approach in the "clean" site situation is to segment the data at a single-well, and ask the question,

> "Has water quality degraded in the last (year)?"

[1] The current RCRA regulations advocate the use of tests of central tendency even at clean sites; however, they recommend tests based only on recent data (for example, the last four independent samples).

A test of trend on only the last four monitoring dates may support this idea. Also, the historic data could be compared to the recent year's data using a test of central tendency.

The ability of these tests to detect groundwater contamination quickly (i.e. when applied quarterly with detection within one or two quarters after a "leak" occurs) depends on the choice of test and other factors which may include the length of the clean water quality record, the variability in the clean data, and the magnitude of the increase in concentration due to contamination.

Also of interest is whether or not specific standards have been exceeded. From a policy perspective, the intended interpretation of a fixed standard must be made clear. Possible approaches include:

- no data should exceed the limit with consideration given to sampling and laboratory error, or
- the historic mean concentration at a well should not exceed this limit[2]; or
- the last-year's mean concentration should not exceed the limit; or
- 95 percent of the population must be below the standard; or
- other.

In the first situation, only sample values close to the standard are given the "benefit of the doubt." In the second and third cases, confidence limits on the mean (where the standard must be below the lower confidence level) may be appropriate. In the last case tolerance intervals are appropriate (Section 3.3.2).

After stating the right question, the next step is to choose a specific test which answers the question. The test must not only have an appropriate experimental design (i.e. answer the right question) but the implicit assumptions of the test must not be grossly violated. As shown in Chapter 2 groundwater quality data may grossly violate the assumption of normality, even after log-transformation of the data. The debate between

---

[2] This choice is rarely appropriate. It may be meaningful at existing sites where disposal predates monitoring and contamination is historic.

the use of parametric and nonparametric tests hinges on the importance of the assumption of normality.

3.0.2 Nonparametric versus Parametric Tests

Strictly speaking *nonparametric* tests are tests which may be based on an assumed probability distribution but which do not involve its parameters. Those procedures which do not rely on a specific probability model at all are termed *distribution-free* tests. While these terms are not synonymous, procedures of either type are generally known as *nonparametric methods.* Most nonparametric tests are based on the ranks of the data rather than the data themselves. Some "information" is "lost" by using ranks rather than the data themselves.

One of the most appealing advantages of nonparametric methods is that they are less likely to be abused. Disregard for fundamental assumptions is the easiest way to abuse statistics. According to J.D. Gibbons (1985a) "If the assumptions cannot be substantiated or are not even known to the investigator, then the inferences may be less reliable than a judicious opinion, or even an arbitrary guess." Nonparametric methods make fewer and less stringent assumptions about the population than those made in parametric statistics. Usually the underlying population or variable is only assumed to be continuous (symmetry is sometimes assumed too). Note however that nonparametric methods do not eliminate the necessity for collecting independent samples.

Advantages of nonparametric tests include the following:

- Data below the detection limit can be incorporated without adjusting the data.
- Nonparametric tests exist for the median, which for skewed data may be a better estimate of central tendency than the mean.
- They do not require the assumption of normality.
- Transformations are not necessary.
- The tests may have greater power to detect contamination when the distribution is skewed and sample size is small.
- The tests are robust to outlying data, hence editing of the data is not

necessary.

Two objective criteria for measuring performance in hypothesis testing are"power" and "robustness." Power is defined as the probability of rejecting the null hypothesis when in fact it is false. Figure 1-1 and the accompanying discussion explored this idea in our situation where the null hypothesis is generally that water quality is clean. Hence, the "power" of a test may be loosely interpreted as the probability of detecting contamination when in fact it is present. If a test has high power even at small sample size it is termed "efficient." A test is "robust" if inferences based on it remain valid, even when one or more basic assumptions are violated. Unfortunately the most powerful tests are those associated with the most assumptions. Conversely robust tests are by definition those with the weakest assumptions. Nonparametric tests are inherently robust, yet they are often criticized for having low "power" compared to parametric counterparts.

Comparisons of the relative performance of parametric and nonparametric tests on real data are difficult to make because it is hard to quantify the relationship between power and robustness at small sample sizes or when the exact normality assumptions are not met. Comparison studies are usually made by evaluating how much power is lost by using a nonparametric test, when all the parametric assumptions are met and sample size is large. If the loss of power is small, then an investigator who has found parametric assumptions to be invalid, or who is unsure of the validity of assumptions may be confident in the choice of nonparametric techniques. Unfortunately, when it is known that parametric assumptions are violated, the actual power cannot be explicitly determined.

The asymptotic relative efficiency (ARE) is a measure of relative performance of two tests at large sample size. The ARE as discussed here is used to compare nonparametric tests to parametric counterparts when parametric assumptions are met. However, the ARE may be used to compare any two tests with similar hypotheses. For practical purposes, the ARE may be interpreted as a ratio of the sample sizes required for two tests to achieve the same power at the same significance level (i.e. at equal Type I and Type II error rates). For example, an ARE equal to 0.85 may be interpreted such that the nonparametric test with 100 observations is approximately as efficient as the parametric test with 85 (if the assumptions of the parametric test are strictly met, and if sample sizes are large). The ARE of a nonparametric test is the minimum relative efficiency, in that the nonparametric test wil never be less efficient than the calculated ARE implies. This is because the relative efficiency of a nonparametric test will always increase if parametric assumptions are not met.. Table 3-1 lists the ARE between analogous parametric and nonparametric tests. These figures must be interpreted with caution because typically sample size is small for groundwater quality data. Table 3-1 shows that generally parametric tests are more powerful in the case of the normal distribution (ARE <1.0); however for other distributions, such as the uniform or double exponential, the nonparametric tests may be as or more efficient (ARE >1.0).

There is no magical test to detect groundwater contamination. The regulatory issues vary from site to site. Also, the hydrogeology and type of contamination will influence a final decision. In the following sections we consider the three general types of stastistical methods introduced in Section 3.0.1:

(1) tests of central tendency (location);

(2) tests of trend; and

(3) prediction, tolerance, and confidence intervals.

In each subsection emphasis is placed on the situations where the type of test is appropriate. The types of water quality questions these tests can answer are discussed.

3.1 Tests of Central Tendency (Location)

To introduce tests of central tendency, a brief review of applicable water quality literature is presented. The basic theory for one approach recommended by EPA is discussed in Section 3.1.1, followed by application to four sites in Section 3.1.2. The findings of this analysis are summarized in Section 3.1.3.

The mean and median are the most common estimators of central tendency. Tests which compare the mean or median of two or more sets of data are termed "tests of central tendency" or "tests of location." Table 3-2 summarizes tests of central tendency which have been proposed to be used with groundwater quality data. We do not recommend the use of several of the tests in Table 3-2. Helsel (1987) applies parametric test statistics using the ranks of the data rather than the concentration values.

TABLE 3-1 Comparison of Nonparametric to Parametric Tests

(adapted from Gibbons, J.D. 1985a,b, and P. Arzberger 1988)

			Asymptotic Relative Efficiency			
Type of Hypothesis	Name of Nonparametric Test	Analogous Parametric Test	Normal	Uniform	Double Expoential	Continuous Lower Bound
Central tendency						
One sample or paired sample	Wilcoxen signed rank test	Student's t-test	0.955	1.0	—	0.864
Two independent samples	Mann-Whitney-Wilcoxen test	Student's t-test	0.955	1.0	1.5	—
k - independent samples	Kruskal-Wallis test	F-test (one way ANOVA)	0.955	1.0	1.5	0.864
Variance						
Two-independent samples	Siegel-Tukey test	F-test	0.608	0.60	—	—
Association analysis						
Two related samples	Spearman rank correlation or Kendall Tau	Pearson product-moment correlation	0.912	1.0	—	—
k-related samples	Kendall Test	F- test (*)	$\frac{.955\,k}{k+1}$	—	—	—

(*) Randomized blocks ANOVA or balanced incomplete blocks ANOVA

TABLE 3-2 Tests of Central Tendency

(P) Parametric (N) Nonparametric

I. At a single well with the data set split at 1) an arbitrary time or 2) remedial action

TESTS	SELECTED REFERENCES[1]
a) Student's t-test and Cochran's Approximation[2] to Student's t-test (P)	McBean and Rover, 1984
b) Mann-Whitney-Wilcoxen test (N)[2]	Doctor et al, 1985, Florida,1985
c) Student's t-test using ranks instead of concentrations (P)[3]	Helsel, 1987

II. Between two or more wells considering 1) all historic data or 2) using only "recent" data.

a) One-way ANOVA with multiple comparison tests (P)[4]	EPA, October 1988, NCASI, 1985
b) Kruskal-Wallis test with multiple comparison tests (N)[4]	EPA, October 1988, NCASI, 1985
c) Two-way ANOVA (P)	NCASI, 1985
d) Analysis of Covariance (P)[5]	Silver, 1986a,b
e) Two-way ANOVA using ranks instead of concentrations (P)[3]	Helsel, 1987

1 A good general reference is "Groundwater Quality Data Analysis" by NCASI (1985). Tests for central tendency for records with dependent observations include Montgomery and Reckhow (1984) and Lettenmaier (1976)

2 These tests were previously recommended by EPA to compare two or more wells. The current trend is to use Group II methods.

3 We do not believe this method is appropriate for determining compliance with groundwater quality regulations. Helsel adopted Conover and Iman's (1976) procedure which they admit can only be justified empirically. Since true nonparametric procedures are available in the situations being considered we do not recommend this theoretically unsound method.

4 For only two wells the one-way parametric ANOVA reduces to the t-test, and the Kruskal-wallis test reduces to the Mann-Whitney-Wilcoxen test.

5 We do not think this method is appropriate because the test addresses whether downgradient wells as a group are different from upgradient wells. Downgradient wells have different potentials for contamination. While some are clean others may be contaminated. There is no physical justification for grouping downgradient wells.

He adapted Conover and Iman's (1976) procedure which they admit can only be justified empirically. Since true parametric tests are available in the situations being considered and since the properties of the adapted tests have not been formally evaluated, we do not recommend Helsel's recommended procedure. Secondly, Silver (1986b) uses an analysis of covariance, where downgradient wells as a group are compared to upgradient water quality. Since downgradient wells have different potentials for contamination, we feel there is no physical justifcation for grouping downgradient wells, and, therefore, we do not recommend this method either.

EPA previously required that Cochran's Approximation to Student's t-test be applied between pooled background water quality data and each downgradient compliance well. Due to criticism of this procedure (EPA, October 11, 1988; Miller and Kohout, undated; Silver, 1986 and McBean and Rovers, 1984b), EPA currently recommends a parametric one-way analysis of variance (ANOVA) or the nonparametric analog called the Kruskal-Wallis test (EPA, 1988). The remainder of Section 3.1 will focus on these tests.

3.1.1 Background on ANOVA

The null and alternative hypotheses for parametric and nonparametric (one-way) ANOVA are[3]:

$$H_0 : \mu_1 = \mu_2 = \mu_3 ..... = \mu_k$$

$$H_1 : \mu_i \neq \mu_j$$

where k is the total number of wells to be compared and μ is the mean or median of a time series. Only wells screened in similar geologic units should be compared; hence k may be less than the total number of wells. If the ANOVA is significant (typically at α = 0.05) there is evidence that at least two wells are different. Multiple comparison procedures can then be used to find which wells are different. Individual comparisons can be made between each compliance well with pooled upgradient water quality. Because these tests are based on the mean or median, they can only answer the

[3]Statistical hypotheses are introduced in Section 1.1.

question whether the central tendency over time differs between wells. The semantics here are important because, as pointed out in the previous section, we are not always interested in historic water quality comparisons.

The hypotheses above are based on the "shift model." This model assumes that the distributions being compared are the same shape, but their means may be different (shifted). Analytically this assumption is interpreted by assuming the sample groups have equal variance (scale). This assumption is frequently violated when contamination exists (as illustrated in Chapter 4). The problem of comparing the location of two distributions of different shape is referred to as the Behrens-Fisher problem. For the two-sample problem Cochran's approximation to Student's t-test accounts for this.[4] We are not aware of an analogous procedure for parametric or nonparametric ANOVA.[5]

The evolution of EPA's regulations from the two-sample t-test to the multiple sample ANOVA came about because many two-sample comparisons are necessary at a site. For example a site with just n=6 compliance wells tested for k=10 parameters will require n x k = 60 two-sample tests. These 60 tests may be substituted by 10 ANOVA tests. In addition, for each parameter which the ANOVA finds a significant difference 6 multiple comparisons must be done. One concern with this approach is that there may still be high probability for false positive error; that is, detection of contamination when none exists. The issue is that the site-wide significance level, α_{sw}, will be very high for so many comparisons.[6] If the significance level for each comparison is α_e, the site

[4]The two-sample problem is when only two data sets are being compared, or when one data set is split at a certain time, and the two time series are compared.

[5]Statisticians and water quality scientists have evaluated the robustness of the t-test to this assumption and others (Montgomery and Loftis, 1987, Boneau, 1960, McBean et al, 1988). Also, the effect of the Behrens-Fisher problem on the Mann-Whitney-Wilcoxen test has been investigated (Fligner and Policello, 1981, Potthoff, 1963). Fung (1979) found the Mann-Whitney-Wilcoxen test to be fairly robust for long-tailed slightly skewed distributions even for sample sizes as small as 10. Montgomery and Loftis (1983) showed that the t-test is not robust for distributions of different shape. Comparisons between Student's t-test and the Mann-Whitney-Wilcoxen test may be found by Pratt (1964), Blair and Higgins (1980) and Rovers and McBean (1981).

[6]α_{sw} is the site-wide probability for each parameter of finding contamination when none exists. See Figure 1-1.

wide significance level for each parameter is

$$\alpha_{sw} = 1 - (1 - \alpha_e)^n$$

where n is the number of compliance wells. (This equation assumes each test is independent.) As n increases, α_{sw} increases. For example if $\alpha_e = 0.01$ with 6 compliance wells, $\alpha_{sw} = 0.06$, but for 20 compliance wells, $\alpha_{sw} = 0.18$. This means there is an overall 18 percent chance that a false positive error will occur. This is not very protective of the owner/operator of the facility. Furthermore if α_{sw} is very high the owner/operator will have legitimate grounds to argue with the validity of significant results.

In the next section, parametric and nonparametric ANOVA are applied to datasets at three municipal sanitary landfills.

3.1.2 Application of ANOVA: Spatial Variability Analysis

In the following evaluation of groundwater quality three main issues are addressed:

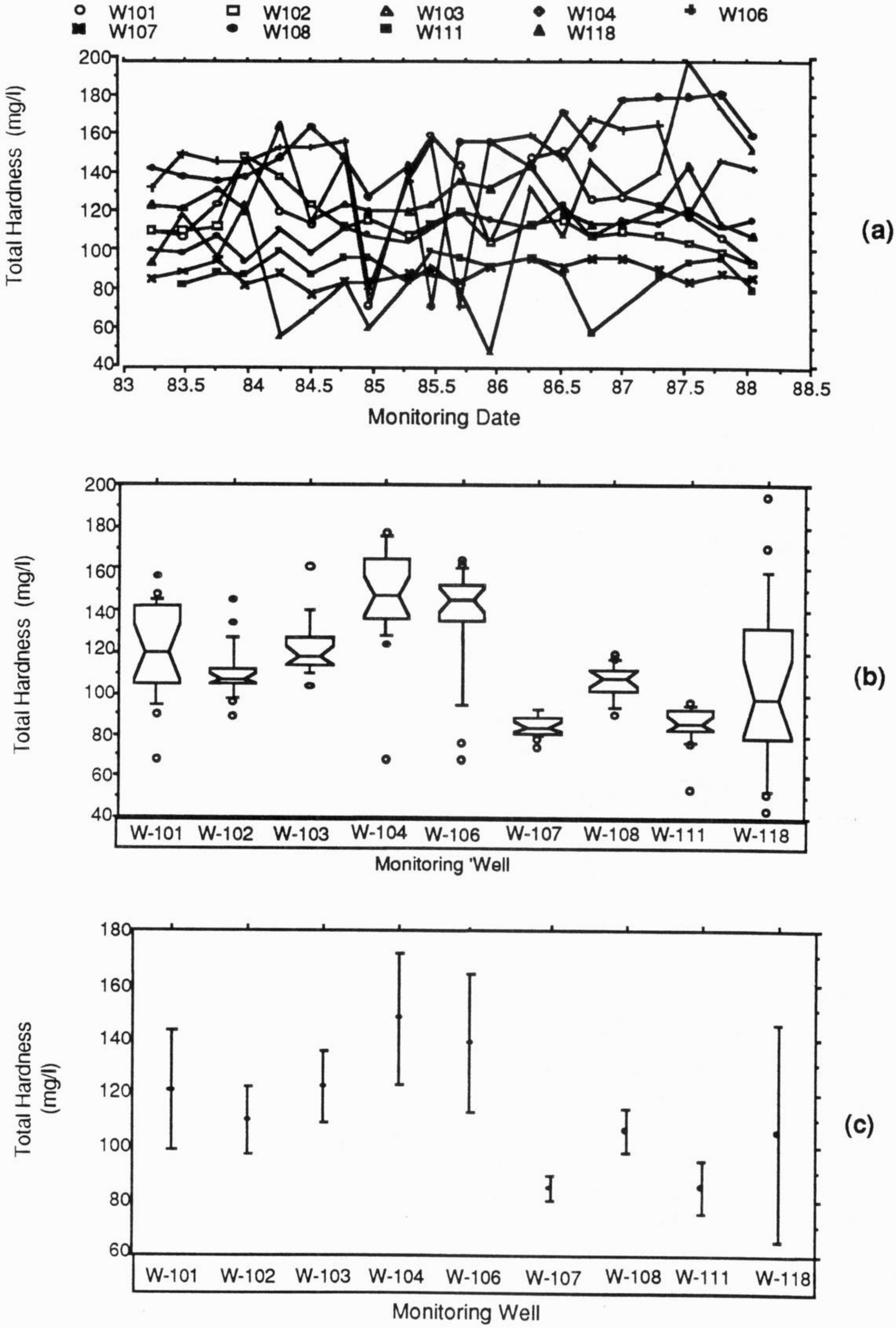
(1) Will natural spatial variations in groundwater quality be detected using ANOVA? If so, natural shifts in mean and shifts due to contamination cannot be distinguished using this method.

(2) Are the assumptions for parametric ANOVA met?

(3) Are results consistent between parametric and nonparametric tests when fundamental assumptions are/are not violated?

The first issue is crucial to the general applicability of this type of test. The second and third issues are raised to consider the performance of these tests under violation of assumptions. These issues are illustrated on Figure 3-4 for total hardness data from wells at the (New) Sauk County landfill. There is no evidence to support groundwater contamination at this site. The confusion evident in Figure 3-4 (a) is clarified in Figures 3-4 (b) and (c). Spatial variation of total hardness is shown by the box plots. The boxes are shifted on the concentration scale; many of the boxes do not overlap. The assumption of equal variance also appears to be violated. The height/length of the boxes are not consistent on Figure 3-4 (b); the range ("length") of the one-standard

FIGURE 3-4 Total Hardness Data from the New Sauk County Landfill



deviation error bars also varies. For example, while well pairs 104/106 and 102/103 appear to have similar variance, the two pairs do not have the same variance. Nor do other wells at the site. Skew may also be visually evaluated on the box plot by seeing if the plots are symmetric and if high outliers are present. These features are evidence of non-normality. The boxplots for W-118 and W-106 clearly show skew.

Three sites were selected for more detailed analysis. These sites are clay-lined facilities with leachate collection systems which are not believed to have contaminated groundwater. "Clean" sites were chosen because we are testing for natural spatial variability only. The sites are:

SITE	LICENSE
• (New) Sauk County Landfill	2978
• Portage County Landfill	2966
• Greidanus Landfill	140

Wells at the Portage County landfill are screened at two levels; thus two separate analyses were performed here. One well at Portage County was deleted from the analysis because contamination was suspected. The only data points deleted from the remaining data sets were those high values very early in the sampling record which may have been a result of well installation and development procedures. The well stabilization period was judged from time versus concentration plots.

A summary of conditions at each site is presented in Table 3-3. Table 1-2 provides additional background information. The materials in which the wells are screened are typical of Wisconsin geology. The analysis was performed for seven water quality parameters. Four parameters are consistently found above detection limits -- pH, specific conductance, total alkalinity and total hardness. Three parameters may be detected at or below the laboratory detection limit -- chloride, chemical oxygen demand and iron.

Three one-way ANOVA tests were performed for each parameter at each site:

TABLE 3-3
Characteristics of facilities chosen for spatial variability analysis

PORATGE COUNTY -- SHALLOW WELLS (12) screened at water table

DNR Well ID No.: Up or side gradient: 1,4,12,14,16,17
Down or side gradient: 9, 23, 24, 26, 28, 30

Length of record: $\tilde{n}$ = 19.24 (geometric mean of n at each well)

Geologic formation: sandy glacial till with cobbles and boulders.

PORTAGE COUNTY -- DEEP WELLS (8)

DNR Well ID No: Up or side gradient: 2,5, 13
Down gradient: 10, 27, 29, 31, 33

Length of record : $\tilde{n}$ = 19.16

Geologic formation: sandy glacial till, coarser than above.

SAUK COUNTY LANDFILL (NEW): WELLS (9)

DNR Well ID No.: Up gradient 101, 102, 103
Down gradient: 104, 106, 107, 108, 11, 118

Length of record: $\tilde{n}$ = 19.76

Geologic formation: sandstone

GREIDANUS LANDFILL: DEEP WELLS IN EXPANSION AREA (3)

DNR Well ID No. : Upgradient 215, 218
Downgradient 225

Length of record: $\tilde{n}$ = 7.92 for Cl and Hardness
$\tilde{n}$ = 8.36 for other parameters.

Geologic formation: glacial outwash -- dense sand with some gravel.

1) parametric ANOVA on raw data;

2) parametric ANOVA on log transformed data; and

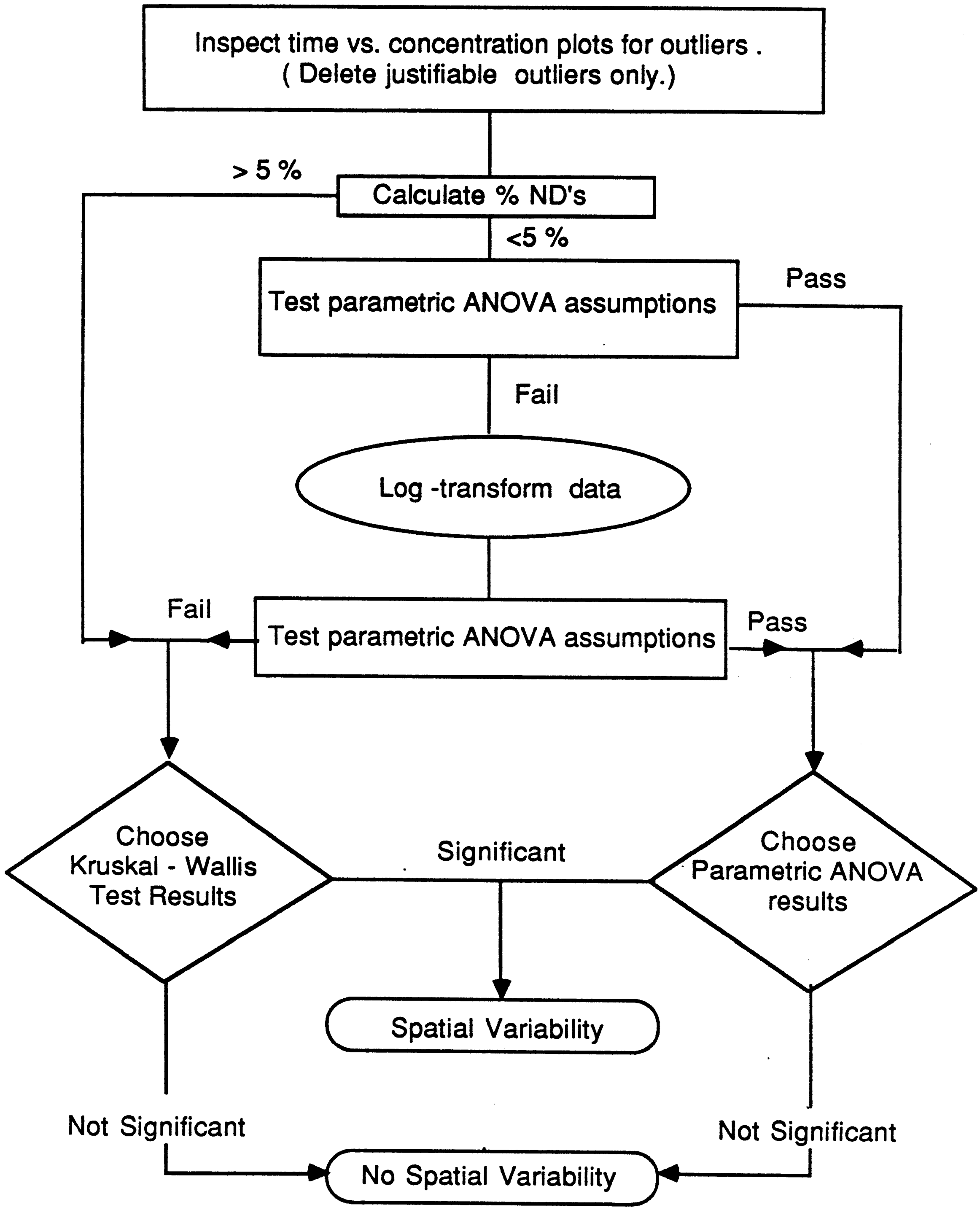
3) nonparametric ANOVA, the Kruskal-Wallis test (KW).

For each parameter (at each site) one of these tests was chosen as the most applicable method based on an evaluation of test assumptions. The preferred test for spatial variability was determined as shown by the flow chart on Figure 3-5. To evaluate the validity of the assumption of constant variance, Bartlett's test for homogeneity of variance was applied to both the raw and log-transformed data. The skewness test was applied grouping all the data at the site after subtracting the individual well means (or means of the logs) prior to analysis. This approach eliminates the effect of shifts in mean (spatial variability) on the overall skew calculation.

The percent of data below detection limits (ND's) was calculated for chloride, iron, and chemical oxygen demand. Except for chloride data at the Greidanus landfill (0 % ND's), the percent ND's exceeded 15 percent of the records site wide. The KW test results are "preferred" for these parameters. "Preferred" is meant in the sense that this test based on the validity of assumptions is most appropriate for this set of data. Four cases had more than 50 percent ND's and even the ANOVA results are suspect.

Table 3-4 summarizes the "preferred" test results. Quite clearly, the KW test is most appropriate for evaluating shifts in mean between landfill monitoring wells. For all three tests which the parametric ANOVA was "preferred," the KW test gave the same result, i.e. significant or not significant at the five percent level.[7]

[7]The choice of the Kruskal-Wallis test is made here by "default." Unfortunately the KW test is also based on the shift model and thus is sensitive to the assumption of equal variance between groups. Technically, the KW test makes the assumption that the distributions are symmetric, which of course is not always the case with water quality data.



Tests of parametric ANOVA assumptions include:

1. Test for homogeneity of variance = Bartlett's test
2. Test for normality : skewness test.
   (applied to site-wide data after removing individual well means)

FIGURE 3-5 Flow Chart for Spatial Variability Analysis

TABLE 3-4
Applicable tests for evaluating spatial variability

PARAMETER	Parametric ANOVA	Parametric ANOVA log-transformed data	Kruskal-Wallis Test
Chloride	1	0	3
Chemical oxygen demand	0	0	4
iron	0	0	4
pH	0	1	3
Alkalinity	0	0	4
Specific conductivity	0	0	4
Total hardness	1	0	3
TOTAL	2	1	25

The parametric ANOVA was rejected for the following reasons:

	Raw Data	Log Data
Presence of ND's	11	11
Violation of variance assumptions	6	3
Violation of normality assumption	1	2
Violation of both assumptions	8	11
Total	26/28	27/28

Note that Bartlett's test rejected the assumption of constant variance 50 percent of the time for both the raw and log-transformed data.

The Kruskal-Wallis test results are presented in Table 3-5. Spatial variability was detected at each site for at least two parameters. The shallow wells at Portage County exhibited significant spatial variability for all parameters. From these results we conclude that natural spatial variability in groundwater may confound the results of ANOVA when applied to detect groundwater contamination. When ANOVA is applied at a site thought to be contaminating groundwater, significant results are not conclusive evidence of contamination.

Significant results may be due to either natural spatial variability or contamination. If ANOVA must be a choice for determination of compliance, 1) Bartlett's test should always be applied, and 2) preliminary ANOVA should be applied at clean wells to see if spatial variability is significant (if there are at least three background wells). If spatial variability is not apparent, then the Kruskal-Wallis test should be applied.

3.1.3 Summary

In conclusion, we have pointed out some severe limitations of tests of central tendency as applied to groundwater quality data. These are:

1) The site-wide significance level, α_{SW}, may be high when many wells are included in ANOVA analyses.

2) Natural spatial variability may be statistically significant. ANOVA results may not be able to discern between natural shifts in mean and those due to contamination.

3) The assumption of homogeneity of variance is frequently violated.

4) Because parametric ANOVA assumptions are found to be invalid in many cases, the nonparametric Kruskal-Wallis test is the preferred test to apply to evaluate spatial variability in groundwater

5) These tests compare the central tendency of the data sets which may not be the right question at sites thought to be "clean."

TABLE 3-5
Results of Kruskal-Wallis Test applied to detect spatial variability

(SIGNIFICANT = Conclude natural spatial variability is detectable.)

PARAMETER	TEST RESULT AT $\alpha = 0.05$		
	Significant	Not significant	p-value
Chloride	2	2	.26,. 25
Chemical oxygen demand	1	3	.66, .55 ,.052
Iron	1	3	.35, .34, .07
pH	3	1	.67
Alkalinity	4	0	-----
Specific conductivity	3	1	.36
Total hardness	3	1	.18
TOTAL	17	11	

3.2 Tests of Trend

Tests of trend can be used to evaluate whether water quality is increasing or decreasing with time. Strictly speaking trend could be observed as either a step function or a gradual increase (usually modelled as a linear function). Step trend should be analyzed using the tests of central tendency discussed in the previous section. Here we are looking at methods to evaluate long-term trend.

Trend tests alone cannot be used to determine compliance with groundwater quality regulations. The tests can only answer the question "Does a positive or negative trend exist?" The tests cannot determine the environmental significance of the trend. The presence of a "small" trend does not mean there is contamination; the absence of trend does not mean there is no contamination. Therefore if a test of trend is used to support the hypothesis of contamination, the results must be linked to exceedanceof standards and to likelihood of contamination.

Tests of trend are also applicable in evaluating the effectiveness of remedial action. However, this type of test should not be used to "predict" when a target concentration will be reached since aquifer restoration is usually not a linear process. A multi-volume document is currently being prepared by EPA on this subject entitled "Statistical Methods for the Attainment of Superfund Cleanup Standards --Draft ."

Table 3-6 lists trend tests proposed in the literature for water quality data with and without seasonal effects. Linear regression analysis is the method most people are familiar with. A least squares regression of the concentration data, y_i, yields a linear best-fit equation

$$\hat{y}_i = m x_i + b$$

where,

$\hat{y}_i$ = predicted mean concentration at time i, the dependent variable

x_i = time, the independent variable

m = the slope of the predicted trend line

b = the y-intercept, a constant.

An f-test on the mean square error of the regression line to the mean square of the

TABLE 3-6 Tests of Trend

(P) Parametric (N) Nonparametric

I. At a single well over 1) entire time series or 2) recent data	SELECTED REFERENCES[1]
a) linear regression with f-test (P)	NCASI 1985 McBean and Rovers 1984
b) Kendall Tau with/without Sen's estimate of slope (N)	Doctor et al 1985 Hirsch et al 1982 Gilbert 1987
c) Spearman's Rho (N)	
d) Seasonal Kendall Tau with/without Sen's seasonal slope estimator (N)	Doctor et al 1985 Hirsch et al 1982 Gilbert 1987
e) Farrell's (Seasonal) aligned-rank test (N)	vanBelle and Hughes 1984

[1] References for trend tests for serially correlated (dependent) data include Hirsch and Slack (1984) and Lettenmaier (1976). Montgomery and Reckhow (1984) present a general methodology for detecting linear trends in lake water quality and recommend specific techniques under various conditions.

"unexplained" error provides a measure of whether the slope, m, is significantly different from zero. Linear regression is very sensitive to outlying values.

Another use of trend tests is to evaluate whether *background* water quality is significantly (gradually) changing in time. In this case, the trend should be removed prior to further analysis (Harris et al, 1987). An apparent trend at a downgradient well cannot be confirmed as evidence of contamination, unless it can be shown that the same trend does not exist in upgradient wells. Detrending is accomplished using the calculated equation from linear regression (above). The predicted mean value of y at time x_i, $\hat{y}_i$, is subtracted from y_i., the observed value of y.

$$z_i = y_i - \hat{y}_i \quad \text{for } i = 1 \text{ to } n$$

z_i (i = 1 to n) are the detrended concentrations. Further analysis on z_i could include tests of central tendency. The linear trend observed in background data is also removed from the compliance well data before analysis.

The nonparametric analogs to the linear regression f-test are Kendall's Tau statistic and Spearman's (Rho) rank correlation coefficient. Usually Kendall's Tau is chosen for water quality data because the test statistic approaches normality at smaller sample sizes than Spearman's Rho (Montgomery et al, 1987). Kendall's Tau is a number between -1 (perfect negative correlation) and 1 (perfect positive correlation). The test statistic, τ, basically evaluates whether the ranks of the data increase with time. (See Gibbons, J.D. (1985b) for a simple derivation of τ.) The calculated value of τ is compared to tabulated values to determine if trend exists (See Appendix A). Because the test is based on the ranks of the data, Kendall's Tau is robust to data outliers.

Linear regression is quite powerful, but analysts tend to delete outlying values without physical justification to get a "good fit." Also, some users will wrongly try to make predictions of "when concentration will return to normal" or "when a standard will be exceeded." The DNR should be aware of the common misuses of regression. When facility reports are submitted to DNR containing linear regression analysis, reviewers should make sure that deletion of data is "physically" justified. Also any predictions made with the regression line should be interpreted as no more than a best guess.

3.3 Confidence, Tolerance and Prediction Intervals

Statistical intervals are used to "bracket" background water quality. Measurements may be compared to the upper bound of the interval to determine contamination. Both the upper and lower bounds are considered for parameters such as pH which may increase or decrease depending on the type of contamination. Each of the techniques are used to answer different questions.

<u>Confidence limits</u> on the mean define an interval within which the true mean of the population will fall (90, 95, 99 percent) percent of the time.

<u>Tolerance limits</u> define a range within which some proportion of the population will fall (90, 95, 99 percent) of the time. Usually this proportion is also 90, 95 or 99 percent.

<u>Prediction limits</u> define an interval within which it can be stated that the next k measurements will fall (90, 95, 99 percent) of the time.

Hahn (1970) explains the difference between these limits.

> A typical astronaut, who has been assigned to a specific number of flights, is generally not very interested in what will happen on the average in the population of all space flights, of which his happens to be a random sample (confidence interval on the mean), or even what will happen in at least 99 percent of such flights (tolerance interval). His main concern is the worst that will happen in the (next) one, three or five flights in which he will personally be involved (prediction interval).

All three (parametric) intervals are symmetric and are calculated based on the models

$$\bar{x} \pm ks \text{ (two-sided)} \quad \text{or} \quad \bar{x} + ks \text{ (one-sided)}$$

where k is a constant obtained from tabulated values. The environmental meaningfullness of these statistics depends on the validity of the assumptions of normality, stationarity and independence.

These limits cannot be used interchangeably. A common mistake is to use confidence limits when tolerance intervals or prediction limits will answer the question of concern. Table 3-7 lists intervals proposed in the literature with application to water quality problems. Table 3-8 lists questions asked from a regulatory perspective and the appropriate method(s) in each case.

In theory question 6 of Table 3-8 is not applicable for determining compliance with groundwater quality regulations. This is because we are not interested in comparisons to average water quality, but rather on comparison of compliance well data to the population of background data. This point is often misunderstood. In fact, conversations with DNR personnel revealed that confidence limits on the mean are currently used for determination of compliance with RCRA regulations at some Wisconsin hazardous waste sites (Tusler, 1988). With the advent of EPA's new rules at these sites, we suggest that alternative procedures be considered (see Section 3.3.1).

In the next two sections intervals are 1) compared to PAL's as estimates of an upper limit of background water quality, and 2) discussed as methods to determine standard exceedances.

3.3.1 Comparison of Intervals to PAL's

How should background water quality be defined? What is a reasonable number above which we suspect groundwater is contaminated? Tolerance intervals, prediction intervals and PAL's have been proposed as estimates of this level. In this section these three estimates are compared. The choice must be made between defining background on a well by well basis or on a site-wide (aggregating data from different wells). In the analysis of four sites presented in Section 3.2, it was found that spatial variability in groundwater quality was common. Both the mean and variance were shown to vary among clean wells. These results suggest that background water quality be defined on a well-specific basis. DNR has adapted this policy for setting PAL's at most sites. However, in some cases site-wide PAL's are in effect. The site-wide approach simplifies the methodology and minimizes the time to calculate PAL's; however, spatial variation is not distinquished from temporal effects. The resulting PAL's may be too high at some wells and too low at others. When spatial variation is present (which we believe to be the usual case) the well-specific approach is preferred.

TABLE 3-7 Confidence, Tolerance and Prediction Intervals

	SELECTED REFERENCES [1]
I. CONFIDENCE INTERVALS 1) to determine standard exceedances, and 2) to determine limits on mean background water quality.	
Normal and lognormal distribution	Gilbert 1987 NCASI 1985
Nonparametric	EPA 1988
II. TOLERANCE INTERVALS 1) to set standards, 2) to determine standard exceedances, and 3) to define an interval within which background concentrations will fall with high probability.	
Normal and Lognormal distribution	Loftis et al 1987 EPA 1988
Nonparametric [2]	EPA 1988
III. PREDICTION INTERVALS 1) to define background concentration interval within which future measurements from downgradient wells are likely to fall.	
Normal and lognormal distribution	Gibbons 1987 EPA 1988 Hahn 1970a, b Hahn and Nelson 1973

1 A good general reference is "Understanding Statistical Intervals" by Hahn (1970).

2 This method requires such a large number of data points to provide a reasonable interval that it appears to be impractical in this application.

TABLE 3-8 Application of intervals to regulatory questions

QUESTIONS	METHOD
1. What is a reasonable upper limit for background water quality?	Tolerance
2. Are downgradient concentrations outside the allowable range of background water quality?	Tolerance Prediction
3. Do new measurements at downgradient wells come from the background population?	Prediction
4. Has a standard been exceeded based on average water quality over a time period?	Confidence
5. Has a standard been exceeded more than a specified percent of the time?	Tolerance
6. Within what range can we state the mean/median of background water quality falls?	Confidence

A PAL equal to $\bar{x} + 3\,s$ may be statistically interpreted as an estimate of the 99.87 quantile of the normal distribution. That is, k=3 is the z-score associated with F(x) = 0.9987 where F(x) is the normal cumulative distribution function. If a time series is independent, stationary, distributed approximately normal, and sample size is large, then it may be stated with confidence that water quality will exceed the PAL less than 1 percent of the time (the exact value is 0.13 percent). At small sample size the probability of exceeding a PAL is not controlled.

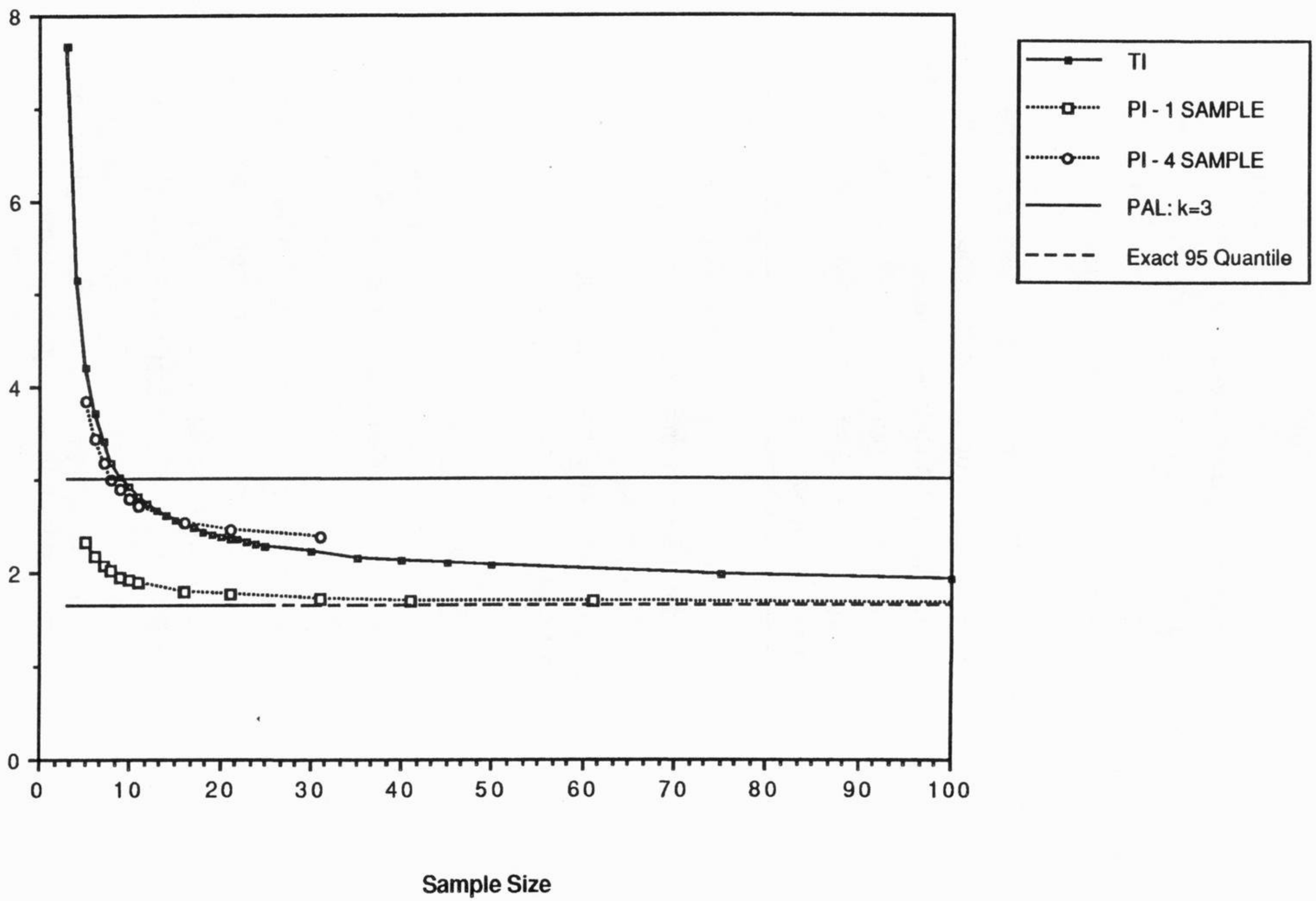
A tolerance interval (TI) , like the PAL, is also associated with a quantile point. However, each end of the TI is an (outside) confidence limit on the exact value of the quantile point. Thus, the TI is a "hedged" estimate of a quantile. For the same quantile point, the (upper) TI will always be larger than the PAL at small sample size, and will approach the PAL at large sample size. Similar to the PAL, the probability of exceeding a TI is not controlled at small sample size.

The prediction interval (PI) does control the probability of exceedance, accounting for both. natural variability and small sample size. Thus, in principle, the PI is the logical choice for an upper background water quality level, although, like the PAL and TI, it depends critically on distributional assumptions. The prediction interval may be calculated for one or more new samples.

At large sample size, when distributional assumptions are met, the TI and PI will estimate the same concentration limit for a given exceedance level. This concept is illustrated on Figure 3-6 for an exceedance probability of 0.95. EPA recommends this significance level for TI and PI calculations (EPA, 1988). The y-axis label "k" refers to the multiplication factor in $\bar{x} + k\,s$ for each type of interval. Prediction interval k's are plotted for one new sample (quarterly comparisons at MSWLF's) and for four new samples (annual comparisons). It is assumed that only one measurement is made per quarter. The PAL (k=3) is also plotted. Figure 3-6 shows the following.

- The PAL is a higher estimate of background water quality than either the PI or TI at sample sizes greater than 10.
- At small sample size, the TI (N<10) and four-sample prediction interval (N<9) exceed the PAL.

FIGURE 3-6 Comparison of Intervals at the 95 % Significance Level



- The TI and one-sample PI asymptotically approach the 95 th quanitle. Note that these intervals would approach the PAL if a significance level of 0.0013 had been used rather than 0.05.

- At small sample size, the TI increases significantly with decreasing n.

- The four-sample PI is greater than the one-sample PI because when four comparisons are made each individual comparison is made at a significance level of (05/4) = 0.125.[8] Thus the four-sample prediction limit asymptotically approaches the 0.9875 quantile point (k = 2.24).

It is important to emphasize that these limits may not be environmentally meaningful if distributional assumptions are not met. The effect of violating distributional assumptions is briefly discussed with an example below.

Table 3-9 presents calculations of the PAL, TI, PI- 1 sample and PI - 4 sample for alkalinity data at DNR Well 18 from the Greidanus Landfill. The intervals are calculated for (1) the normal distribution and (2) the lognormal distribution. At this well, alkalinity varies over a wide range (minimum 93 mg/l; maximum 465 mg/l). Yet there is no reason to suspect contamination or any grounds to delete high or low data. Comparing the skewness coefficient for the two distributions implies that the normal distribution better represents these data (since γ is closer to 0). Sample size is small (9), thus the tolerance interval is greater than the PAL, as mentioned above. Inspection of the resulting limits reveals that the lognormal intervals are unreasonably high. The normal calculations appear to be much better estimates of upper limits of background water quality in this case.

The above discussion has illustrated different methods for establishing an upper limit for background water quality. Important conclusions are summarized below.

- The prediction interval in theory directly answers the right question: What is the concentration associated with an allowable exceedance probability given the natural variability in the data and the sample size?

- The PAL is a conservative (high) estimate of background water quality compared to the more sophisticated TI and PI, except at small sample size. (when the TI and PI are calcuated at a significance level of 0.05 as recommended by EPA).

[8]That is, when using the Bonferroni multiple comparison procedure.

TABLE 3-9
Example Calculation of Statistical Intervals

Alkalinity data from Well # 18 Greidanus Landfill							
Distribution	$\bar{x}$	s	γ	PAL k=3.0	TI k = 3.03	PI (1) k=1.96	PI (4) k=2.89
Normal	259	105	0.334	574	577	465	562
Lognormal	5.47	0.465	-0.784	959	972	591	909

TI = Tolerance interval
PI (1) = One sample prediction interval
PI (4) = Four sample prediction interval

- The TI may be unreasonably high at very small sample size (<8).
- If distributional assumptions are not met, all of the methods may yield results which are not environmentally meaningful.

These three methods all predict some upper limit for background water quality above which contamination is suspected. The environmental significance of the exact number, however, is unknown.

The PAL is simplest to calculate. The tolerance interval is almost as easy, except instead of using "3" as the multiplier in $\bar{x} + ks$, a value is obtained from a table. The value of the tolerance interval k is less than 3 at large sample size and greater than 3 at small sample size. The prediction interval is most difficult to calculate and is dependent on the number of future monitoring rounds the interval will cover. All three of these methods assume the data are normally distributed which we know is not always true. The methods may also be applied to the log-transformed data if necessary. We do not recommend aggregating data from different wells to set any of these limits.

Because the PAL method is simplest and yields a conservative (from the industry perspective), but not unreasonably high, upper limit of background water quality, we recommend that DNR continue to employ the PAL "as is" at waste disposal sites. If facility owners are dissatisfied with PAL values, DNR should suggest that they calculate prediction intervals as described in EPA (1988) for each well. This procedure is also presented in Appendix A. At hazardous waste sites, we recommend that prediction intervals be calculated for each well rather than the t-distribution confidence intervals currently employed. These recommendations are synthesized into flow charts in Chapter 5.

The DNR currently sets PAL's at existing sites with known contamination as well as at new facilities. The PAL at a contaminated well is set based on a clean well(s) screened in a similar stratigraphic unit. We do not believe that this practice is appropriate or correct, due to spatial variations in the mean and standard deviation of water quality in groundwater. Therefore, if a site is known to have contaminated groundwater, PAL's may not be necessary at all, since PAL's from a regulatory perspective are intended to be early warnings of groundwater contamination. PAL's may be appropriate at a contaminated facility as clean-up goals for a remedial action procedure. For existing sites with contamination, we suggest that the procedures discussed in Sections 5.2 and 5.3 be used to build a defensible case of contamination.

In determining whether a PAL has been exceeded, DNR may be confident that contamination exists when more than one sample at a well exceeds a PAL, particularly if the PAL is based on a reasonable sample size (8 or more independent samples at a well) and no data are deleted from the time series.

3.3.2 Intervals to Determine Standard Exceedances

This section addresses the use of intervals to determine exceedance of externally defined standards such as EPA's maximum contaminant levels, Wisconsin's enforcement standards, and PAL's set based on a percent of these standards. PAL's set based on background water quality are addressed in the previous section.

As discussed in Section 3.0.1 the definition of a standard must be clear. Interpretation of federal and state regulations is not straight forward. NR140 implies that ES's and PAL's set based on the ES are not-to-be-exceeded limits. However, if it can be shown to the agency "that a scientifically valid determination cannot be made that the preventive action limit or enforcement standard ..... has been attained or exceeded based on consideration of sampling procedures or laboratory precision and accuracy...." then no remedial response shall be required (Wisconsin DNR, 1988). The discussion of these regulations will therefore focus on using intervals to determine exceedances with consideration of 1) sampling procedures, and 2) laboratory error. Unlike the WI regulations, the EPA hazardous waste regulations (as interpreted in the Draft Guidance Manual) imply that some samples may exceed the standard. EPA recommended procedures are discussed following the discussion of NR 140.

NR 140. A strict interpretation of NR 140 would be to consider each exceedance "real" unless the possibility exists that sampling or laboratory error were the cause of the high value. Under current regulations, if an exceedance occurs which the agency or owner feels is an anomalous high value due to sampling procedures or gross laboratory error, the well must be resampled. Gross laboratory error implies data transcription error, sample mislabeling, etc.and must be distinquished from true laboratory error which is associated with the precision and accuracy of the actual analysis.

If a sample and the corresponding "re-sample" exceed the standard, the only "unaccounted" reason for a false positive exceedance is true laboratory error. If an exceedance is near the standard, it may be that the "true" sample concentration does not exceed the standard. Tables 3-10 summarizes laboratory accuracy confidence ranges at the Wisconsin State Lab of Hygiene for several parameters (Songzoni, 1988). These levels have been calculated for many other compounds as well. Accuracy confidence ranges are applicable for judging sample exceedances close to the standard. The accuracy confidence range defines an interval within which the true sample concentration will fall 95 percent of the time. These levels are determined from spiked samples; that is, samples with a known concentration prepared by the laboratory. The confidence range is expressed as a percent of the true concentration. For example, if a total hardness concentration value of 403 mg/l was measured, the interval of 392 mg/l (0.973 X 403) to 411 mg/l (1.02 X 403) will contain the true sample concentration 95 percent of the time. If a PAL of 400 mg/l is in effect, the possibility exists that the true alkalinity concentration may be below the standard by 8 mg/l. Therefore, the standard may not have truly been exceeded.

As one would expect, Table 3-10 shows that only a small percent of the total concentration should be due to analytic error. These estimates may be lower than at other laboratories which do not employ as strict quality control procedures. Rice, Brinkman and Muller (1988) reported on a quality assurance program for groundwater samples which evaluated eight laboratories for precision and accuracy. They concluded that the reliability of laboratory analyses should not be taken for granted. (They also found analytic reliability to be independent of the prices charged by the laboratory.)

In summary the strict interpretation of NR 140 presented here allows for resampling when sampling procedures or gross laboratory error is suspected. Otherwise, sample values are considered true exceedances except when laboratory accuracy ranges indicate that the true sample concentration may be below the standard. EPA recommends a less stringent approach for evaluating quarterly groundwater reports from hazardous waste sites.

TABLE 3-10
Accuracy Confidence Range for Non-RCRA Samples;[1]
Determined from Spiked Samples at WI State Lab of Hygiene

PARAMETER	METHOD	95 % CONFIDENCE RANGE (%)	99 % CONFIDENCE RANGE (%)
Hardness	200.1	97.3-102	96.0-103
COD, LL	280.2	90.4-108	86.0-113
COD, ML	270.2	93.1-105	89.0-108
Sulfate	370.2	94.0-107	91.0-110
Iron, Flame	500.1	93.3-104	90.0-107
Chloride	140.2	97.3-103	96.0-104

[1] Non-RCRA samples include surface and drinking waters, groundwater, and domestic and industrial wastes.

RCRA Subtitle C and D Discussion.

The federal hazardous waste regulations (RCRA Subtitle C) require 16 independent samples annually. Four measurements are to be collected each quarter but must be independent samples. To achieve independence the quarterly data may be collected daily, weekly or monthly depending on groundwater flow. (EPA,. October 11, 1988). At waste sites governed under Subtitle D of RCRA, sampling consists of one sample four times a year. EPA recommends two methods for determining exceedance of health or welfare standards: confidence limits and tolerance limits.

The first method EPA (1988) discusses is construction of a 99 percent confidence limit on the mean of the most recent four measurements. The standard is then compared to the lower limit of the confidence interval. In the hazardous waste case, the mean of the four quarterly measurements are used to construct the confidence interval. Thus, in this case, the confidence interval method addresses the question

"Does the mean of the quarterly sample exceed the standard?"

This approach "allows" some samples to exceed the standard. It is possible that three out of four quarterly measurements could exceed the standard, yet this approach may not "detect" an exceedance. This possibility exists because the confidence interval is very sensitive to the standard deviation of the four measurements; data drawn from a contaminated regime is likely to be quite variable, since contaminated groundwater is not well mixed. Because data from a contaminated regime are most likely not from a single population, we do not believe that distributional parameters should be calculated using these data. Furthermore, confidence intervals based on such a few samples are always wide, since information is limited.

While the confidence interval could be applied to the MSWLF situation, it must be recognized that when only four samples are collected per year the question becomes,

"Does annual average water quality exceed the standard?"

At some sites this question may not be adequate. Contamination by highly toxic substances may require quicker action. More frequent sampling at MSWLF's could alleviate this problem. Yet, even if 16 samples were collected, like at HWS's, the high variability of groundwater may cause unreasonably wide confidence intervals.

EPA(1988) also presents a tolerance interval method. This approach would be applicable when a permit is written specifying that a standard is not to be exceeded more than a specifed fraction of the time. EPA suggests that the four quarterly measurements be used to construct an upper tolerance limit. If this limit is below the standard then the site remains in compliance. It is important to note that the choice of an upper tolerance limit is much more protective of the environment than the lower confidence limit discussed above. We do not like this procedure for the same reasons given for the confidence interval approach. Data from a contaminated distribution should never be used to directly estimate distribution parameters.

The strict interpretation of an exceedance as discussed with respect to NR 140 is the most environmentally sound approach to comparing data to standards. When contamination exists, the confidence interval and tolerance interval approaches recommended by EPA will rely on estimating the mean and standard deviation with data which may not be from a single population. Furthermore, the high variability typically observed in contaminated time series will cause a very low confidence limit to be calculated.

3.4 Summary and Regulatory Perspective

In this chapter three types of statistical tests were discussed: tests of central tendency, tests of trend and statistical intervals. The discussion emphasized the questions addressed by each type of test addresses. Both parametric and nonparametric tests were considered. The evaluation of statistical methods yields some insight into how DNR may effectively determine compliance with NR 140 at waste disposal sites. On the other hand, several of EPA recommended procedures for use at RCRA hazardous waste sites were shown to be ineffective.

As evidenced by the Wausau Paper Mills example in Section 3.0.1, many statistical questions may apply at a site. At sites with historically evident groundwater contamination, no statistical tests may be necessary, since a strong case for contamination may be built solely using information on site hydrogeology, groundwater flow and water quality graphs. PAL's are intended to act as flags of potential water quality degradation. We do not feel they are applicable at existing sites with contamination unless they are calculated as a target value set as a remedial action goal.

There is no magical test to conclusively detect groundwater contamination. Some statistical tests may confuse the issue of documenting groundwater contamination. Violation of statistical assumptions may lead to erroneous Type I and Type II error rates. Inexperienced analysts may apply the wrong test to answer the right question or vice versa. The evaluation of statistical tests presented here will hopefully contribute to solving this problem.

Tests of central tendency both parametric and nonparametric were shown to have severe limitations. ANOVA tests were applied to three state-of-the-art sanitary landfills which are not suspected of contaminating groundwater. The analysis showed that natural spatial variability may be statistically significant. <u>ANOVA results may not be able to discern between natural variations in mean and those due to contamination</u>. Also, the parametric assumption of normality and the assumption of constant variance were found to be frequently violated for both raw and log-transformed data sets. The nonparametric Kruskal-Wallis test was found to be the preferred test of central tendency, yet we recommend its use only for testing for spatial variability, not for detecting contamination. Because spatial variability was frequently observed, we recommend the use of well-specific comparisons, such as statistical intervals and trend tests for detecting contamination.

Tests of trend may be used as supporting evidence of contamination. These tests however cannot be used alone since the sole presence of a trend is not conclusive evidence of contamination. However, <u>a strong positive trend of magnitude great enough to cause standard exceedances is powerful evidence of contamination.</u> We recommend use of Kendall's Tau, a non-parametric correllation coeffiecient, rather than linear regression techniques. This recommendation is made because linear regression may be easily misused and is dramatically affected by data outliers.

Confidence, prediction, and tolerance intervals were discussed from two perspectives: 1) as methods for establishing upper limits for background water quality (i.e. lower limits of potential contamination), and 2) as tools for determining standard exceedances. As mentioned above, we feel intervals set based on background water quality are the preferrred method for statistically detecting groundwater contamination. The intervals should always be calculated on a well specific basis rather than on a site-wide basis since the mean and variance at different wells are likely to vary significantly. A comparison between tolerance intervals, prediction intervals and PAL's showed that PAL's consistently are the highest estimator of background water quality, except when sample size is small. These results however were dependent on the choice of a 95 percent significance level for the tolerance and prediction interval. This choice was made based on EPA recommendation. The prediction interval is the most theoretically sound approach to setting background levels, since it is a concentration value associated with a specified exceedance probability and takes into account sample size and natural variability.

We recommend that the DNR continue to rely on the PAL as an early warning of groundwater degradation. PAL's should always be set on a well-specific basis rather than site wide since spatial variability is believed to be common. From a statistical perspective the prediction interval is superior to the PAL, since this method is theoretically correct when the implicit assumptions are met. If a facility owner is dissatisfied with DNR's calculated PAL values, prediction intervals are recommended as an alternative approach.

At hazardous waste sites where quarterly comparisons are made between background water quality and compliance well data by facility owners/operators, we suggest that the prediction interval approach be considered before other methods. The interval may be updated, perhaps annually, to give a better, probably lower, estimate of background water quality. The updated interval will probably be lower because at the same exceedance level, the interval decreases with increased sample size

When data are to be compared to health and welfare standards, we recommend that DNR take a strict approach to determine exceedances. Resampling should always be allowed. However, all results should be considered "true" unless it can be shown that there is gross error due to sampling procedures or laboratory error. In all other cases, exceedances should be given the benefit of the doubt only if laboratory accuracy ranges indicate that the true sample concentration may be below the standard. This will only be the case when sample values are very close to the standard.

We do not support the use of confidence or tolerance intervals for determining standard exceedances as recommended by EPA. These methods, and in particular the confidence interval, are considerably less protective of the environment than the method outlined above. Our primary concern with these methods is that they require estimation of the parameters for a (possibly) contaminated distribution. These estimates of the mean and standard deviation are meaningless since samples are probably not from the same population, i.e. leachate plumes are not homogenous mixtures of pure chemicals. Furthermore, the estimates of the standard deviation may be greatly inflated, since groundwater contamination is highly variable.

The information presented in this chapter implies that in a court of law, statistics applied to groundwater quality data may be easily challenged. Fundamental assumptions of all these methods are frequently violated. A strong case built on hydrogeology, disposal history, and water quality graphs may be supplemented with statistical test results. However, a groundwater contamination case which rests heavily on statistical conclusions will never stand up to detailed examination. Statistics should be viewed as an admittedly imperfect regulatory tool used to confirm apparent contamination for determination of compliance with groundwater regulations.

The recommendations summarized above are synthesized into a methodology in Chapter 5 with specific recommendations for changes to NR140.

CHAPTER FOUR
STATISTICAL SCREENING OF THE GROUNDWATER QUALITY DATABASE

4.0 OVERVIEW

While the first three chapters of this report have focused on determining compliance with groundwater quality regulations, this chapter presents the results of a statistical analysis on Wisconsin's database of water quality data at waste disposal sites. The landfill groundwater database includes over 300 licensed sites, each with a number of wells, and water quality data for an array of constituents at each well.

The DNR is currently working on setting preventive action limits (PAL's) at landfill monitoring wells for various constituents. The PAL (for those constituents without a mandated enforcement standard) is set based on a review of background water quality (see Section 1.2). The current procedure to set PAL's for these indicator parameters requires a thorough review of site history, hydrogeology, and historic water quality data. This procedure, as one might expect, is time consuming.

Hence, the broad objective of our analysis of the groundwater database was to prioritize particular sites for the setting of PAL's and to help minimize the time required to set PAL's for all the sites. In order to meet this goal, several more specific objectives included:

(1) to characterize the data available for each site;

(2) to develop a "predictor" of groundwater quality change with respect to background water quality at a site;

(3) to use this predictor to group landfill sites into categories of similar groundwater impacts.

In March 1988 we submitted a report to DNR entitled Summary Report Task 3 "Statistical Screening of the Groundwater Quality Database." In that report a "predictor" of groundwater quality change with respect to background water quality was developed and applied to the database. Landfill sites were grouped into categories of similar groundwater impact. In subsequent meetings with DNR it became apparent that this predictor did not work well at sites where more than 50 percent of the wells at the site were contaminated. Since this is often the case at sites with significant problems, we continued this line of research. The "predictor" presented in this chapter is simpler than the one which we first proposed. Also, a larger control group of sites was used to evaluate the predictor performance. Therefore, this chapter is intended to replace in entirety our previous report.

We described the groundwater quality data using a database management program developed specifically to meet our needs. For each site a variety of information was generated, including: the number of dates each parameter was tested at each monitoring well, the number and name of parameters tested per well, and the number of replicate samples.

No surface water monitoring stations or other non-well points were considered. A total of 316 sites are included in the database with a total of 4202 wells. If it is assumed that an average of five PAL's will be set at each well, then a total of 21,010 PAL values must be determined. These numbers are approximate since some sites are in the process of closing, and new sites are added on an ongoing basis.

Many environmental parameters are monitored at the landfill sites; however this study focused on the eight chemical parameters most frequently monitored at landfills:

- chloride
- chemical oxygen demand
- sulfate
- iron
- pH
- specific conductance
- total hardness
- total alkalinity

In the site characterization it was found that 9 sites had no data for these parameters, and 79 sites have less than 8 monitoring dates for all parameters at all wells. We did not address these sites in the statistical analysis. The remaining 228 sites were screened for evidence of contamination.

4.1 Development of Predictor of Groundwater Quality Change

The general strategy employed to develop a groundwater predictor was, first, to evaluate in detail the site conditions for a subset of 20 sites from the groundwater quality database. Then this subset was used to develop and test a predictor. To evaluate whether groundwater contamination had occurred, water quality data from well samples was reviewed in conjunction with information on disposal history, waste type and hydrogeology. In this study, we analyzed a control group of 274 wells from 20 landfill sites in Wisconsin for evidence of contamination. A summary of the characteristics of these sites was presented in Section 1.3. Not only was geology, groundwater flow, site history and waste type considered, but the water quality data at each well were tested for increasing trends in time, and also compared to other wells. The 274 wells were grouped into four categories based on the evidence that contamination is present or absent. The four categories are:

Group I	Presumptive evidence that well is clean
Group II	Evidence that well is probably clean
Group III	Evidence that well is probably contaminated
Group IV	Presumptive evidence that well is contaminated

The procedure used to group wells is similar to the flow charts presented in Section 5.1 for setting PAL's based on background water quality. More specifically,

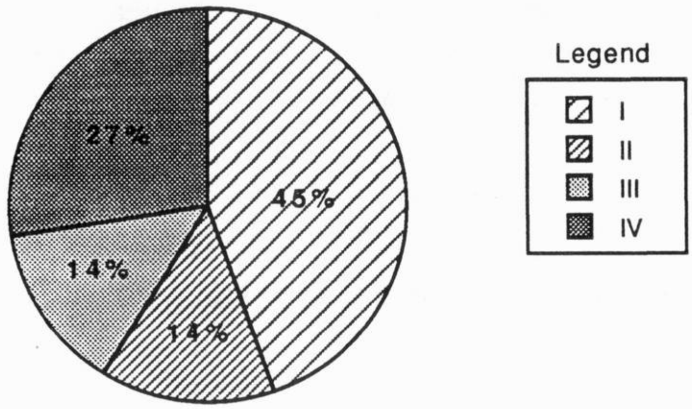
- We met with DNR personnel to discuss background information at each site.
- Information on site history, geology and groundwater flow was summarized.
- Wells with sufficient data were identified and located on a site map.
- Time versus concentration plots and box plots were constructed for each parameter for wells screened in similar geologic formations.
- Kendall's Tau two-sided test for trend was applied. Significant trends were

noted on box plots.

- Inferences drawn from the parameter plots for each well were summarized.
- Wells were grouped into categories I, II, III, or IV.

A summary of the wells considered and their associated group is presented in Appendix B. Group I and II wells may be used to determine background water quality. Figure 4-1 shows that of the 274 control group wells, 59 percent are either Group I or II (clean), while 41 percent are in Group III and IV (dirty).[1]

FIGURE 4-1 Breakdown of Control Wells into Four Water Quality Groups



Each dataset for the control wells was statistically summarized including:

$\bar{x}$	the sample mean
s	the sample standard deviation
x_{ln}	the sample mean of the log-transformed data
s_{ln}	the sample standard deviation of the log-transformed data
m	the median

[1] Note that the 99 control group wells evaluated in the previous study (Goodman & Potter, 1987) were re-classified based on two years of additional data. Some well groups were changed: most changes were "downgrading" wells from I and II to III and IV.

IQR the interquartile range[2]

These summary data were explored for use as possible predictors. In theory, a perfect prediction will always correctly assign wells as "clean" or "dirty." In practice, some group I and II wells will be predicted "dirty," and some group III and IV wells will be predicted "clean." Figure 4-2 summarizes these errors for a predictor on the control group wells. Clearly we want to maximize the success rate, and minimize the probability of false positive and false negative error.

FIGURE 4-2 Illustration of Errors Inherent in the Contamination Predictor

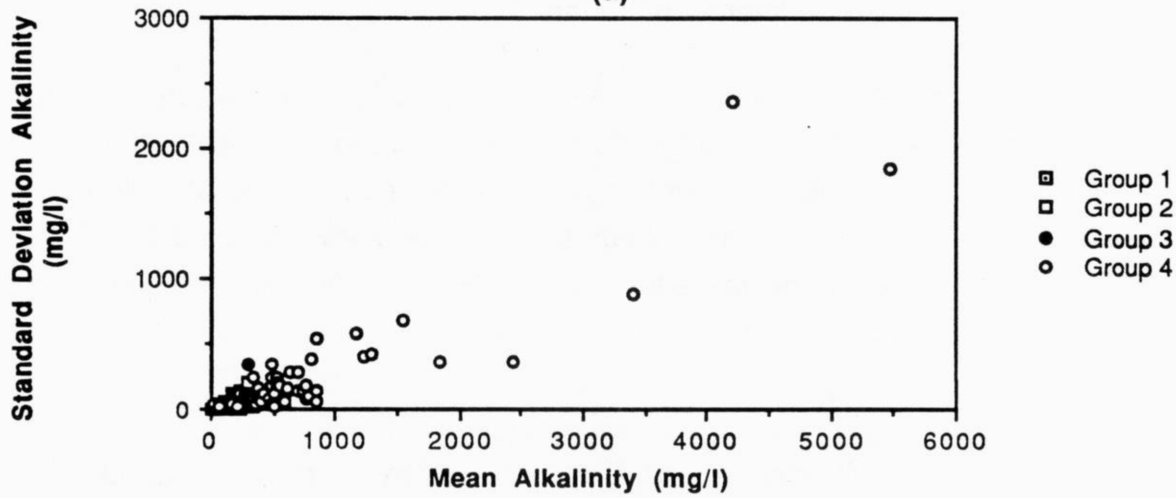
		Predicted Situation: Clean	Predicted Situation: Dirty
True Situation	Clean	Group I & II	Group I & II False Positive Error Rate, α
True Situation	Dirty	Group II & IV False Negative Error Rate, $1-\beta$	Group III & IV Dirty Success Rate = β

Preliminary analyses included plots of mean versus standard deviation and median versus fourth spread as shown on Figure 4-3 (a), (b) and (c) for alkalinity data sets. Of these three plots, the lognormal mean and standard deviation plotted on Figure 4-3 (b) most clearly separate Group I and II from Groups III and IV. Clearly wells with very high contamination make plots of the mean versus standard deviation (a) and median versus

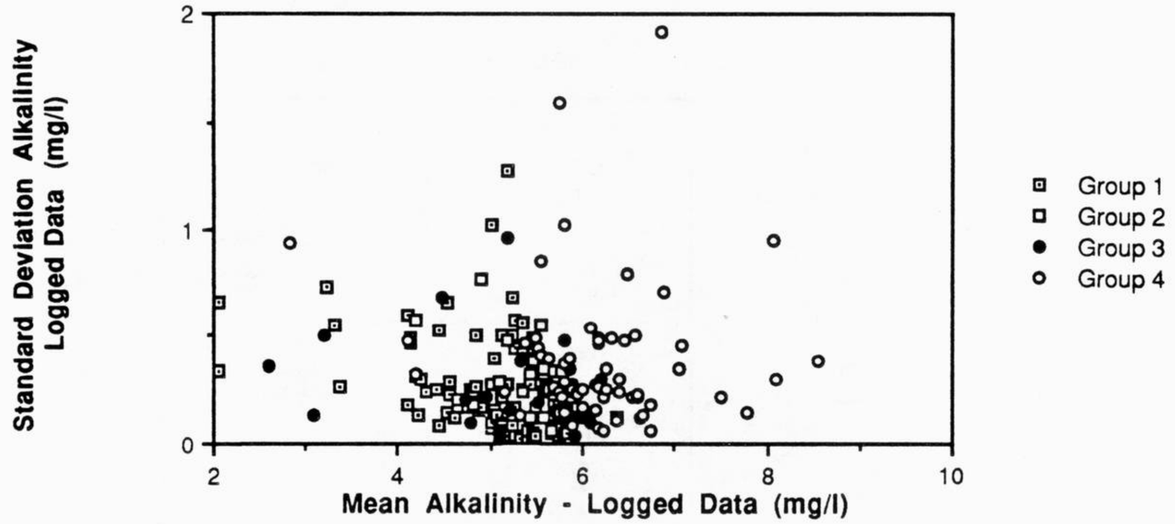
[2] The IQR is usually referred to as the fourth spread when estimated from a sample.

FIGURE 4-3 Exploratory Data Analysis of Possible Predictor Statistics

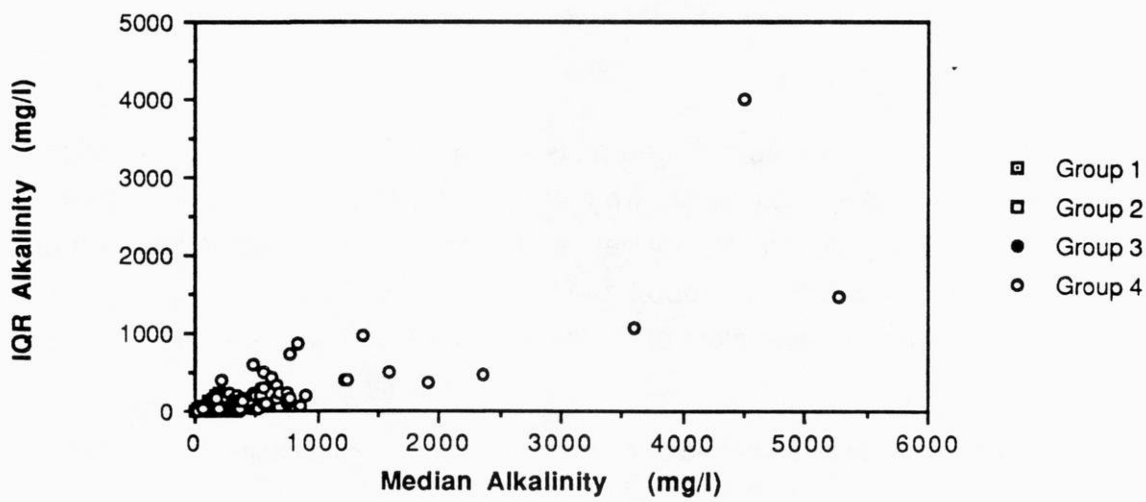
(a)



(b)



(c)



IQR (c) difficult to interpret. These two plots are replotted on a logarithmic scale on Figure 4-4 (a) and (b). [3] These plots illustrate that data from "dirty" wells (squares) have significantly higher variability and central tendency than data from "clean" wells (circles).

Further inspection of Figures 4-3 (b) and 4-4 (a) and (b) reveals that:

- The lognormal standard deviation is consistently low for clean wells (Figure 4-3 (b)) Since the standard deviation is sensitive to outlying values, this result is not surprising.
- Both Figure 4-4 plots show a clear difference between clean and dirty wells, which is the objective for a predictor of groundwater contamination. The statistics are consistently low for group I and II wells, relative to Group III and IV.
- Less "overlap" of clean and dirty wells is apparent in Figure 4-4 (b) than Figure 4-4 (a).

While this third point is difficult to see on the graphs, this finding implies that the median and IQR are better statistics for delineating contamination than the mean and standard deviation. Inspection of similar plots for all parameters confirms this fact. Fewer false positive results (prediction of clean wells to be dirty) will occur when the predictor is based on the nonparametric estimators.

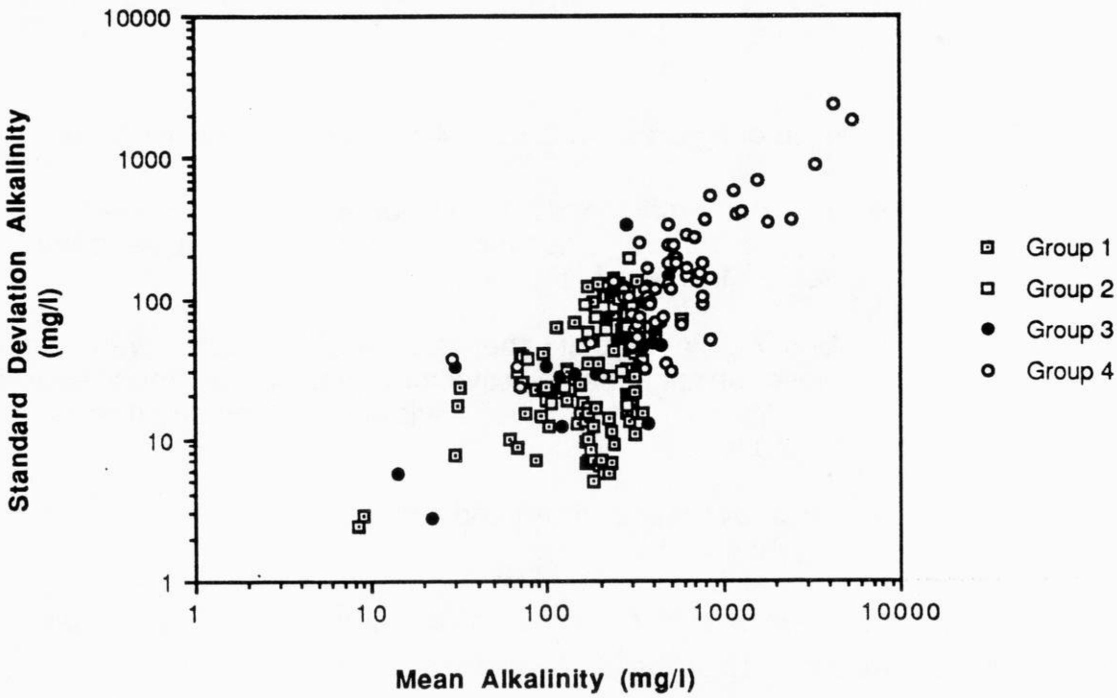
Considering Figure 4-4 (b) , we can define a region within which all, or close to all, the clean wells fall. A limit of 350 and 150 mg/l for the median and IQR respectively appear to capture most of the Group I and II wells. Any data points above these limits could then be predicted as contaminated. This is the basic idea of our predictor.

Two of the eight parameters do not successfully separate the well groups. Figure 4-5 (a) shows, the median and IQR for pH data cannot be successfully used to separate groups, hence this parameter was dropped from the analysis. Also the plot for sulfate, Figure 4-5 (b), shows fewer wells are sampled for this parameter. Also many of the dirty wells

[3] Note that Figure 4-3 (b) is not the same as Figure 4-4 (a). The former is a plot of the mean and standard deviation of the log transformed data sets, while the latter is a graph of the statistics calculated for the raw data, but plotted on a log scale for clarity purposes.

FIGURE 4-4 Further Data Analysis of Possible Predictor Statistics

(a)



(b)

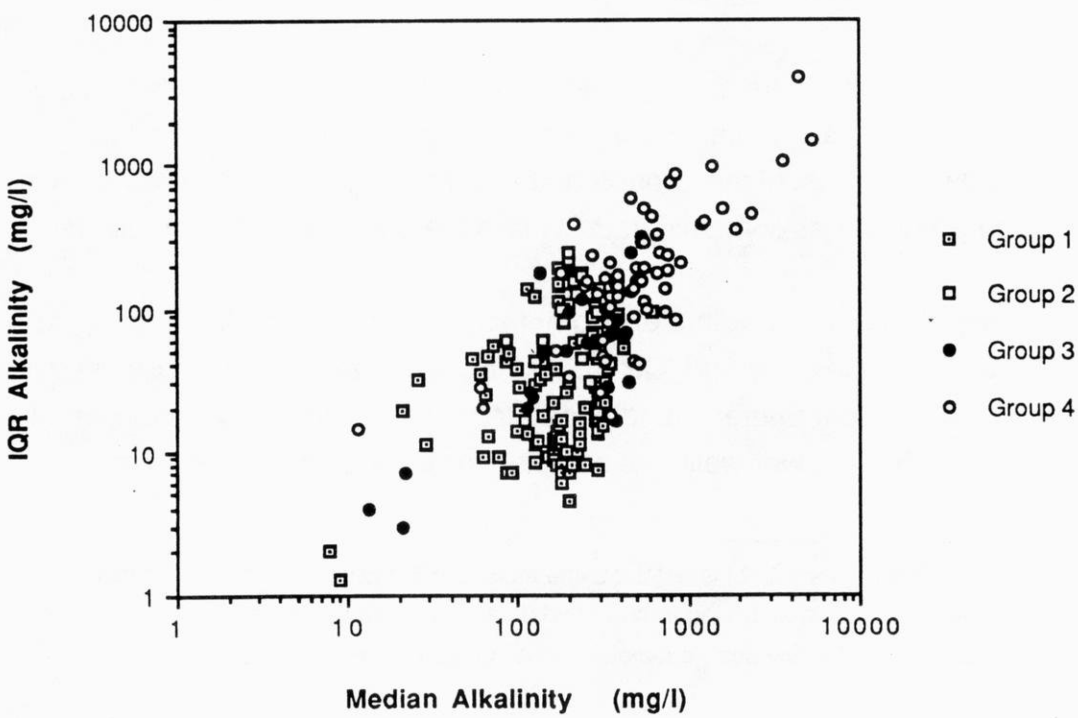
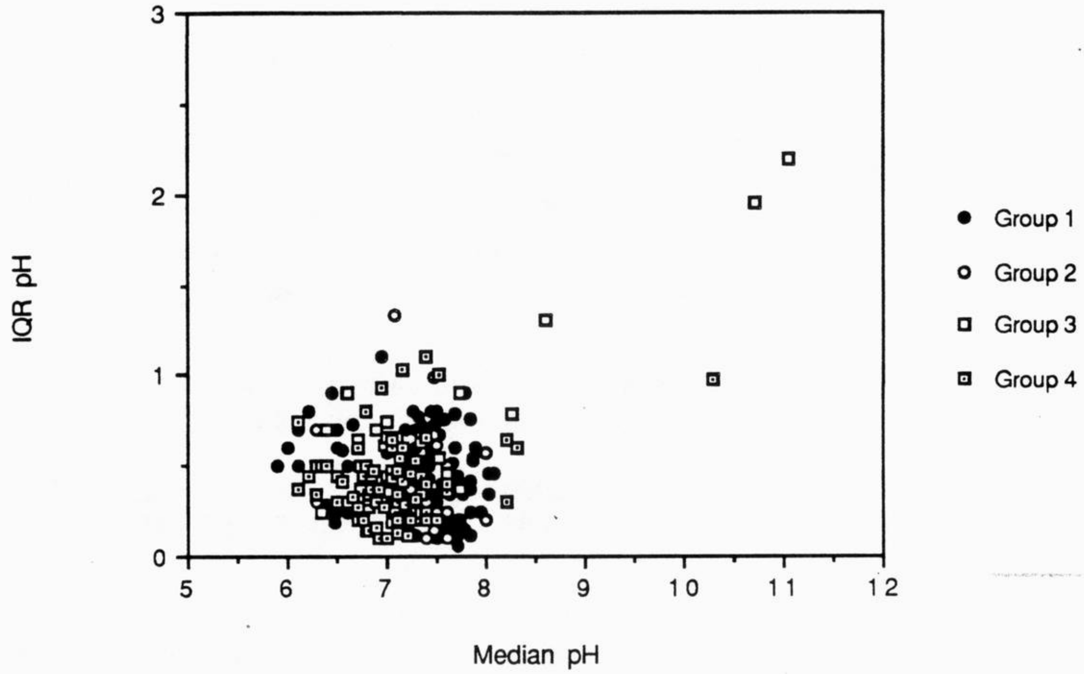
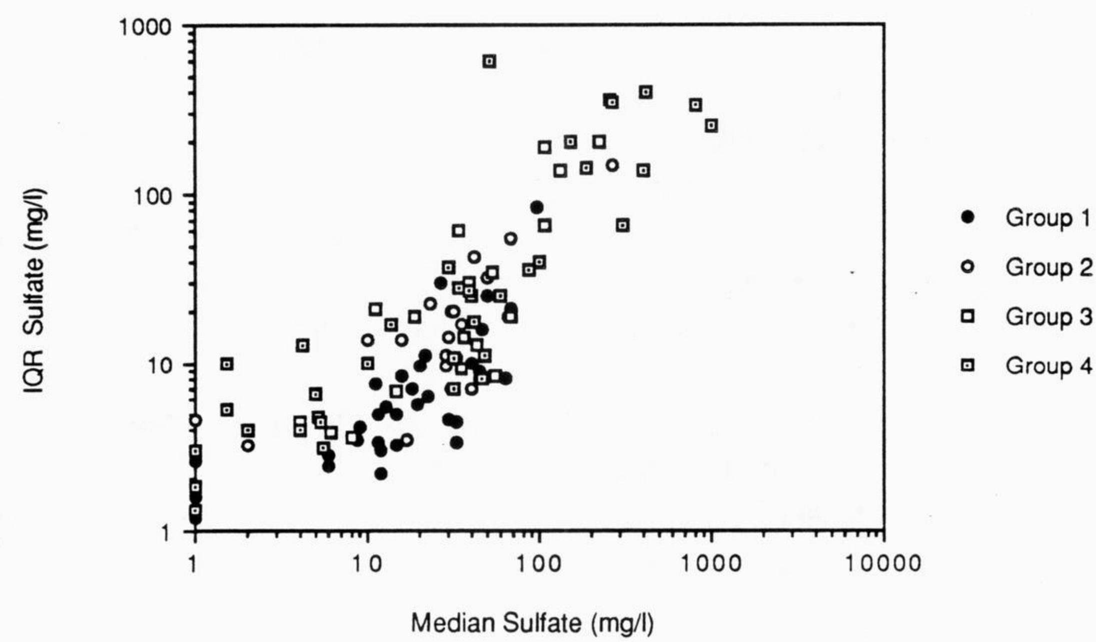


FIGURE 4-5 Parameters which do not successfully separate well groups

(a)



(b)



have low median and will not be predicted "dirty." This high false negative rate does not help our objectives, therefore sulfate was excluded in the final predictor analysis.

Concentration limits were estimated for each of the six remaining parameters below which almost all of the clean well data falls, and above which contamination appears likely. For each parameter some "dirty" wells will be predicted to be clean, since not all parameters may be elevated at a particular site. But, if we summarize the wells predicted "dirty" over <u>all</u> parameters we will most likely capture a high percent of dirty wells. Thus, two possible predictors considered were:

> (1) if for <u>one</u> parameter at a well the median <u>or</u> the IQR fall above the established limits, then the well is predicted "dirty," and
>
> (2) if for <u>two</u> parameters at a well the median <u>or</u> the IQR is above the established limits, then the well is predicted "dirty."

The two predictors were applied to the control group using limits visually determined from the median/IQR log-scale plots . A computer program was written to calculate false positive and false negative rates for any concentration limits. By adjusting the limits and re-running the program we tried to optimize the success rate and minimize the false negative rate of the predictor.

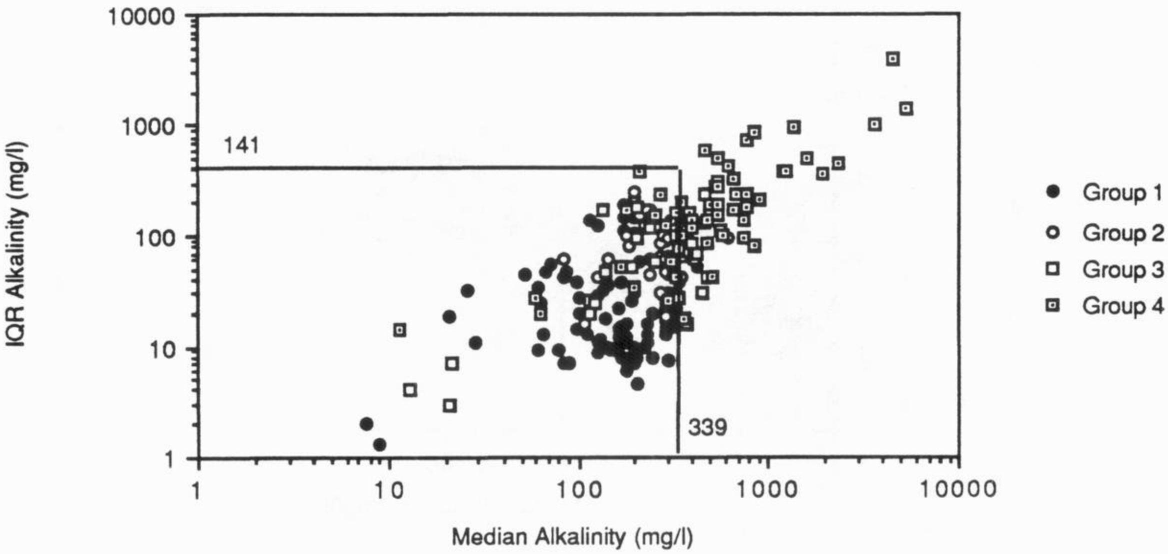
It was found that the <u>one</u> parameter predictor had an 86 percent success rate but a 22 percent false positive rate (clean wells predicted dirty). The <u>two</u> parameter predictor reduced the false positive rate to 6 percent, but also reduced the success rate to 76 percent. We then inspected the false positive and negative results and made adustments to the concentration limits. Figures 4-5 (a), (b), (c), (d), (e), and (f) show the prediction limits for each parameter which were found to maximize the success rate and minimize the false positive rate. The concentration limits are listed in Table 4-1.

The final rule was, if a well had <u>two</u> parameters with median or IQR above the prediction limits, then the well was predicted dirty. This predictor has an estimated success rate of 84 percent and an estimated false positive rate of 9 percent based on analysis of the 274 control group wells.

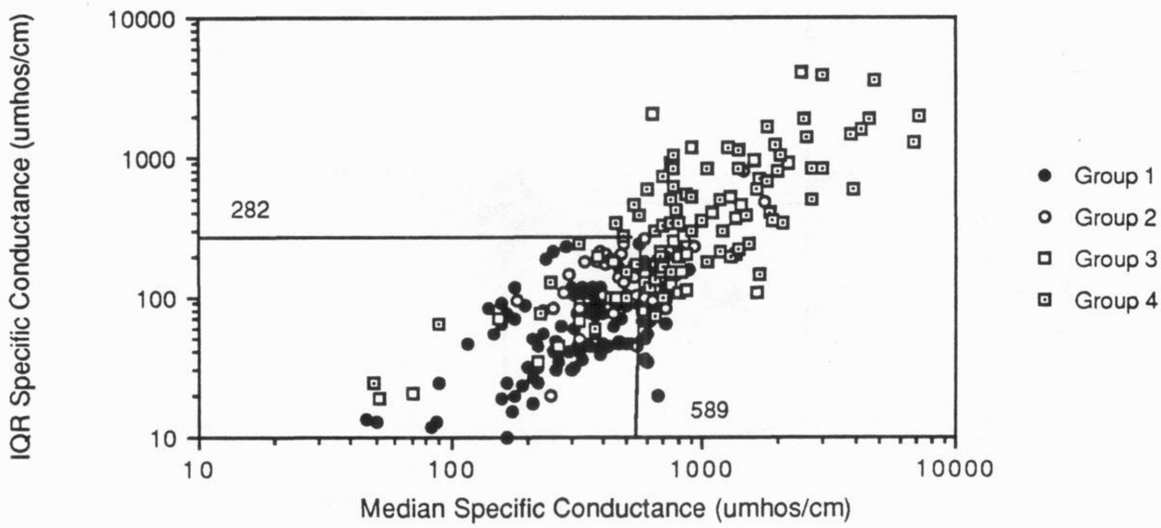
An analogy may be drawn between our prediction limit for each parameter and the PAL

FIGURE 4-6 Median & IQR Prediction Limits for Six Parameters

(a)



(b)



(c)

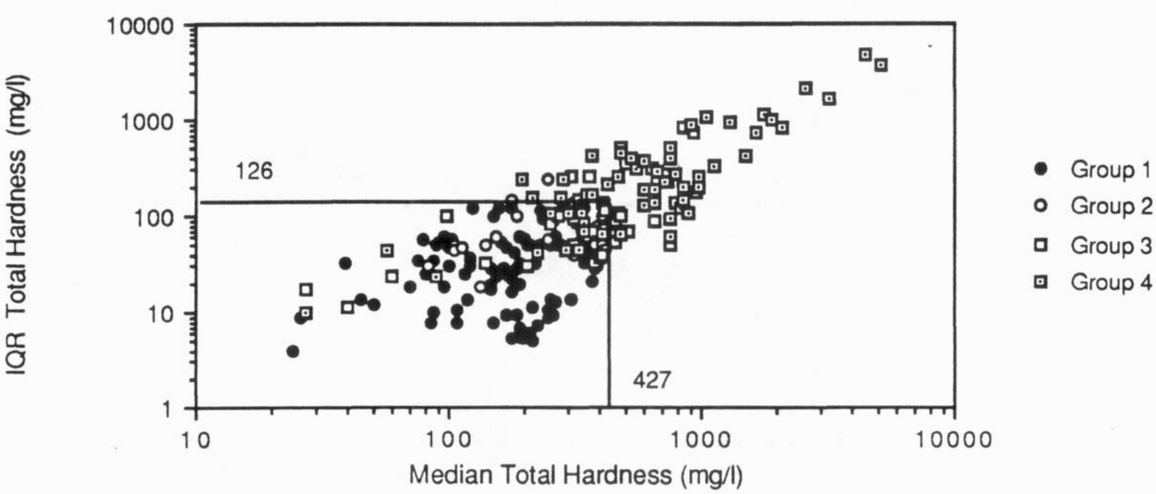
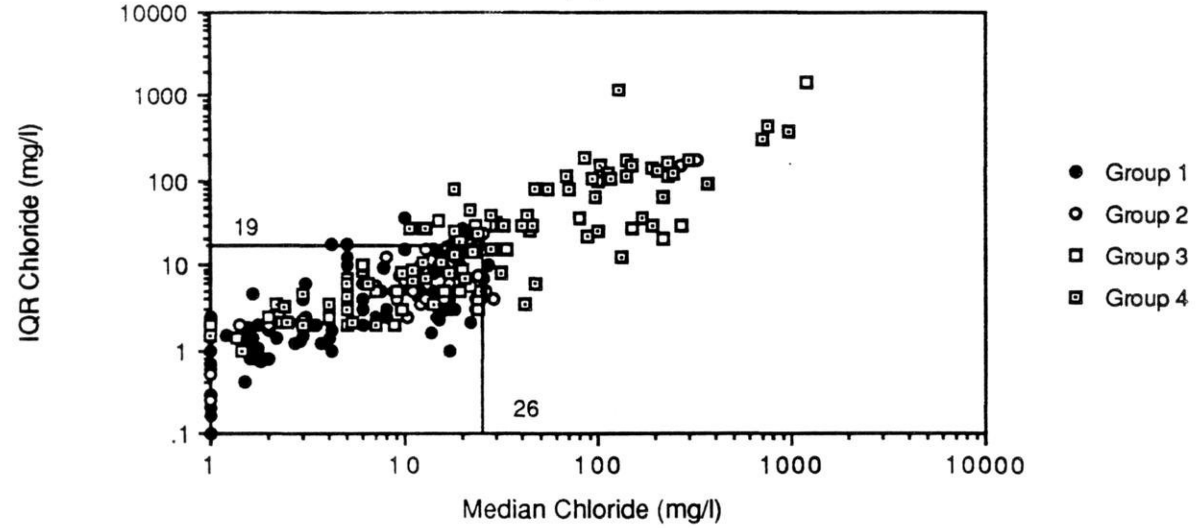
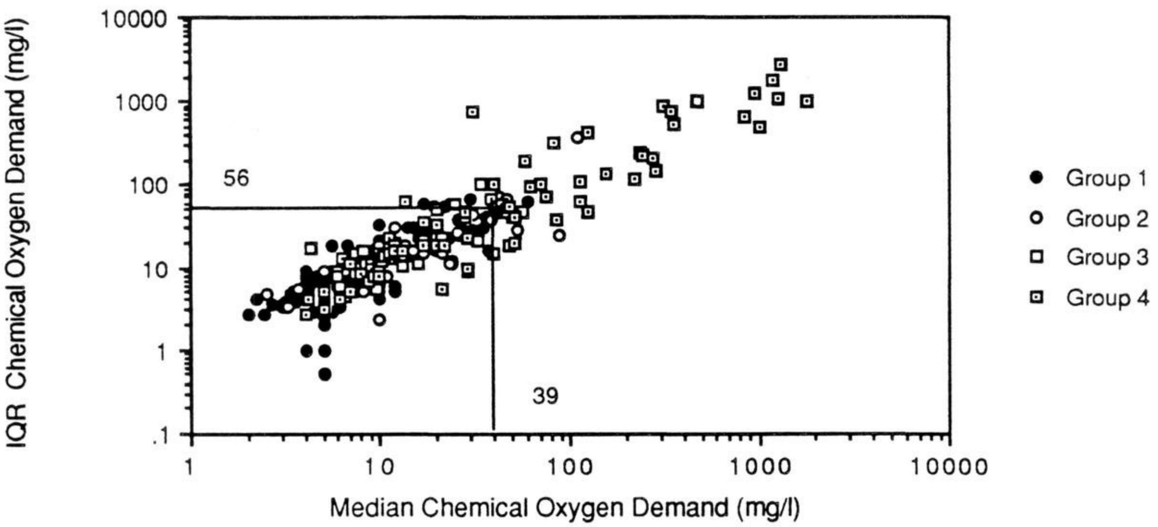


FIGURE 4-6 (Continued)

(d)



(e)



(f)

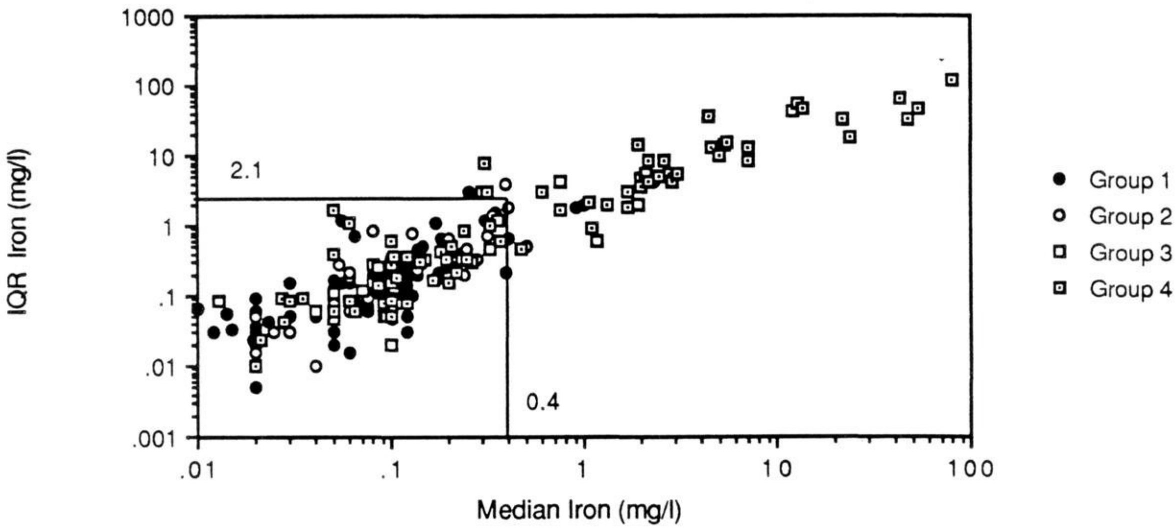


TABLE 4-1
Concentration limits used for groundwater contamination predictor

PARAMETER	MEDIAN (mg/l)	IQR (mg/l)
Alkalinity	339	141
Specific Conductance [1]	589	282
Total Hardness	427	126
Chloride	26	19
Chemical Oxygen Demand	39	56
Iron	0.40	2.10

1 Units are μmhos/cm

(for parameters without an enforcement standard). The prediction limits were chosen to be above background water quality at the 20 control sites, and hence may be thought of as upper limits of "clean" water quality. Both the PAL and the prediction limits are intended to act as flags of background water quality.

A comparison was made between final PAL's set for ten sites (five of which were included in the analysis) and the prediction limits. Since the predictor is based on either ther median or the IQR exceeding a prediction limit, a full comparison cannot be made. PAL values were only compared to the concentration limit set for the median. PAL's which are mandated as 50 percent of the enforcement standard were not included (i.e. PAL's for chloride and iron). However, PAL's established as a minimum increase over background values are included. Of the 277 PAL values considered for specific conductance, COD, hardness and alkalinity, only 74 exceeded the prediction limits. 47 of the 74 PALs which exceeded the prediction limits were for COD, and 23 of these 47 PAL's were within 10 mg/l of the limit.

This comparison shows that the prediction limits individually are conservative estimates of an upper limit for background water quality. Yet, at the 20 sites considered in this study, 85 percent of the "dirty" wells are still detected using the overall predictor.

4.2 Screening of the Database

The chosen predictor was applied to all wells in the database which had at least eight monitoring dates for one or more parameters. As mentioned previously, a "dirty" well is defined as having two parameters exceeding the prediction limits.

A simple scheme was used to categorize landfill sites into groups of similar groundwater quality impacts. Two criteria were used: (1) the number of wells at the site; and (2) the percent of "dirty" wells at the site.

A site with a large number of wells is likely to be a large site, or a site at which there is a recognized problem, or a site which is close to a community or valuable resource.

The choice of the second criteria is based on the assumptions that:

(1) there are more downgradient wells than upgradient; and

(2) the more dirty wells, the higher possibility of adverse environmental impact.

Note that the local extent of contamination is considered using these grouping criteria.

Table 4-2 summarizes the site grouping results. As shown, 16 groups were defined based on the two criteria. The number of wells at a site (criteria 1) was divided into 4 categories: 1-5 wells, 6-10 wells, 11-20 wells and greater than 20 wells. The percent of wells, with sufficient data, predicted dirty (criteria 2) was also divided into 4 groups as shown on Table 4-1. 24 sites with more than 10 wells are found to have over 75 percent of the wells (with enough data) predicted dirty. 41 sites are found to have all or almost all clean wells (i.e.29+5+7+0). Table 4-3 lists these clean sites and their license numbers. Table 4-4 lists the 24 sites (16+8) with greater than 10 wells with over 75 percent or more predicted dirty. The complete results are in Appendix C.

TABLE 4-2 Summary of Site Grouping Results

Percent of wells predicted dirty	Total number of wells: 1 - 5	6 - 10	11 - 20	> 20	TOTAL
0 %	29	5	7	0	41
1 - 50 %	16	10	19	8	53
50 - 75 %	22	26	8	9	65
> 75 %	29	16	16	8	69
TOTAL	96	57	50	25	228

NOTE: Percent of wells predicted dirty calculated based only on wells with enough data. Total number of wells includes wells without enough data.

The sites listed in Table 4-3 most likely are not having a significant impact on groundwater. We feel that the method used in this analysis is powerful at detecting low groundwater impact. The method has limitations in cases of high contamination as discussed below.

The lists presented in Tables 4-3 and 4-4 and Appendix C do not represent an absolute ranking of facilities. Rather the list should be interpreted as groups of sites which appear to have similar groundwater impacts using our method. It is important to recognize that impact is determined based on statistical analysis of only indicator parameters.

TABLE 4-3 Sites predicted to have low groundwater impact

No.	FACILITY NAME	LICENSE
1	BRIDGEPORT LANDFILL	445
2	CNTY EAU CLAIRE-SEVENMILE CRK	2821
3	CNTY JUNEAU	2565
4	CNTY LA CROSSE	2637
5	CNTY MONROE-RIDGEVILLE SITE	2858
6	CNTY ONEIDA	2805
7	CNTY PORTAGE LANDFILL	2966
8	CNTY SAUK SANITARY LANDFILL	2978
9	CTY GALESVILLE	2738
10	CTY NEW RICHMOND	2492
11	CTY PHILLIPS	57
12	CTY SHAWANO - PHASE 2	3069
13	CTY WISCONSIN DELLS	2712
14	EXXON MINERALS COMPANY	2977
15	HUGHES REFUSE & LANDFILL CO	2776
16	JACKSON CNTY SANITARY LF INC C	2004
17	LEADFREE LANDFILL-BRIDGEPORT	2959
18	MERCURY MARINE LANDFILL	2603
19	N.O.W. PAPER CORP. FLY ASH LF	2964
20	NEKOOSA PAPERS INC	2891
21	NEKOOSA PAPERS, INC	2811
22	NORTHRNESTATESKPWR-DEERPCRE	2767
23	PATS STORAGE LAGOON	3003
24	RIVERSIDE SANITATION LANDFILL	738
25	SCOTT PAPER CO LANDFILL	2846
26	TN GRAND RAPIDS	693
27	TN HALLIE	2807
28	TN MENOMONIE	2659
29	TN MINOCQUA-MERCER LAKE SITE	1559
30	TN SHERMAN	2856
31	TN STUBBS-DISTRICT 5 LANDFILL	2909
32	TN SUGAR CAMP-SOUTH SITE	2884
33	VALLEY SANITATION CO, INC	2686
34	WARD PAPER COMPANY LANDFILL	2991
35	WATERFORD SEPTIC SERVICE	2894
36	WAUPACA FOUNDRY, INC	2089
37	WIS ELECTRIC POWER CO-HWY 32	2801
38	WIS ELECTRIC POWER CO-HWY 59	918
39	WIS PUB SERV CORP-WESTON #3 LF	2879
40	WIS PUBLIC SERV CORP-LEGNER	3067
41	YOURCHUCK'S SANITARY LANDFILL	2010

TABLE 4-4 Sites predicted to have high groundwater impact.

No.	FACILITY	LICENSE
1	APPLETON PAPERS, INC	30
2	BARRETT LANDFILL, INC	1940
3	BERGSTROM PAPER LF-NEENAH	2446
4	CNTY FOND DU LAC	2358
5	CNTY KEWAUNEE SW BALEFILL	2975
6	CNTY MILWAUKEE HWY DEPT	881
7	CNTY WINNEBAGO	611
8	CONSOLIDATED PAPERS-KRAFT DIV	1838
9	CONSOLIDATED PAPERS-WQC	2488
10	CTY ASHLAND	177
11	CTY SHAWANO	2342
12	CTY SUPERIOR-WIS POINT LF	12
13	FLAMBEAU PAPER CORP	2756
14	HOLTZ & KRAUSE, INC	674
15	JAMES RIVER NORWALK-NORTHLAND	2893
16	NEKOOSA PAPERS (LIME SLUDGE)	2614
17	SANITARY TRANS & LF-DELAFIELD	719
18	TORK ALUM MUD DISPOSAL SITE	1892
19	TORK LANDFILL CORPORATION	652
20	VULCAN MATERIALS CO	2998
21	WASTE MANAGEMENT OF GREEN BAY	3
22	WASTE MNGT OF WI, INC-POLK	307
23	WASTE MNGT OF WIS, INC-CITY DS	37
24	WASTE MNGT OF WIS-BROOKFIELD	1

Limitations of this analysis include,

(1) Contamination from specific heavy metals, volatile organic hydrocarbons, or other hazardous constituents is not directly evaluated; instead, indicators of water quality change were considered. These indicators are naturally found to occur in low concentrations in groundwater. The result is that some wells with contamination only by a specific contaminant may be overlooked.

(2) False positive results may occur at wells screened in geologic formations which naturally have highly variable background water quality or high background concentration. Glacial till for example may exhibit naturally high water quality variation.

Table 4-3 and 4-4 and Appendix C may be interpreted in several ways:

(1) Sites which have few monitoring wells (<10) with greater than 50 percent of the wells flagged may be sites which need additional monitoring. While each site must be considered individually, our results indicate that contamination is apparent, but information is limited.

(2) Sites with greater than 50 percent of flagged wells should have higher priority for in-depth investigation than other sites.

(3) When targeting a particular site for investigation Appendix C may be used to get a first understanding of water quality at a landfill.

(4) Results should not be interpreted as an absolute ranking of water quality at landfills.

The first item above applies to 93 sites in Wisconsin, that is 41 percent of the sites included in this screening (and 29 percent of the 316 licensed sites-- as of August 1987). These results imply that high priority should be given to expanding the monitoring system design at existing landfills in Wisconsin. Item two suggests that the sites listed in Table 4-4 should be given high priority, however many of these are currently closely monitored. Review of Appendix C may identify sites not currently assigned to personnel which may warrent additional review.

4.3 Characteristics of Background Water Quality

The exploratory data analysis presented in Section 4.1 illustrated that the median and IQR may be used to predict levels above which contamination is likely. This section illustrates characteristics of background water quality in Wisconsin as defined by these same statistics. Data from 161 Group I and II wells were used to generate histograms of the distribution of clean water quality at the 20 landfill sites investigated in this study. Figures 4-7 and 4-8 present these distributions for the median and IQR respectively. The prediction limits are drawn on each figure. Note that the scales on these figures are log base 10, i.e 1= 10 mg/l, 2 = 100 mg/l etc. When interpreting laboratory reports of sample values, an analyst may wonder if a value is really high/low with respect to other sites in the State. Figure 4-7 may be used by new DNR hydrogeologists to get a feel for what typical concentrations for background water quality are.

Figure 4-7 may also be used when evaluating contamination at a site. The standardization procedure currently used at DNR to allow many parameters to be plotted on one figure employs both the median and interquartile range (See Goodman and Potter for a discussion of this procedure). The standardization procedure is useful in this respect, however at a site where there is contamination at more than 50 percent of the wells the values for the standardization statistics will not be representative of background water quality (i.e. the zero point of the graph will not represent background conditions). The values of the site-wide median and median fourth spread may be compared to the appropriate graphs on Figure 4-7 and 4-8 to get a feel for the reasonableness of the values.

Appendix D contains a statistical summary of these distributions. Also included are summary statistics for the distribution mean, standard deviation, and (standardized) skewness coefficient for the raw data and log-transformed data (not presented here).

4.4 Summary and Conclusion

A statistical analysis of groundwater quality data collected at monitoring wells at landfill sites in Wisconsin was conducted. A characterization of the data available showed that 316 sites with 4,202 wells are included in the database, however only 228 sites were included in the analysis.

A predictor was developed of groundwater quality change with respect to background water quality. Performance of the predictor was evaluated using a control group of 274 wells analyzed in detail in a previous study. The predictor was defined as not-to-be exceeded upper limits of background water quality for each parameter. The chosen predictor required at least two parameters at a well to exceed the prediction limits before the well was predicted "dirty." It is estimated that this predictor successfully observes contamination 84 percent of the time, while having false positive prediction (clean wells estimated dirty) only 9 percent of the time. The percent of wells considered dirty and the total number of wells at a site were then used as criteria to place all the sites into 16 groups as summarized in Appendix C.

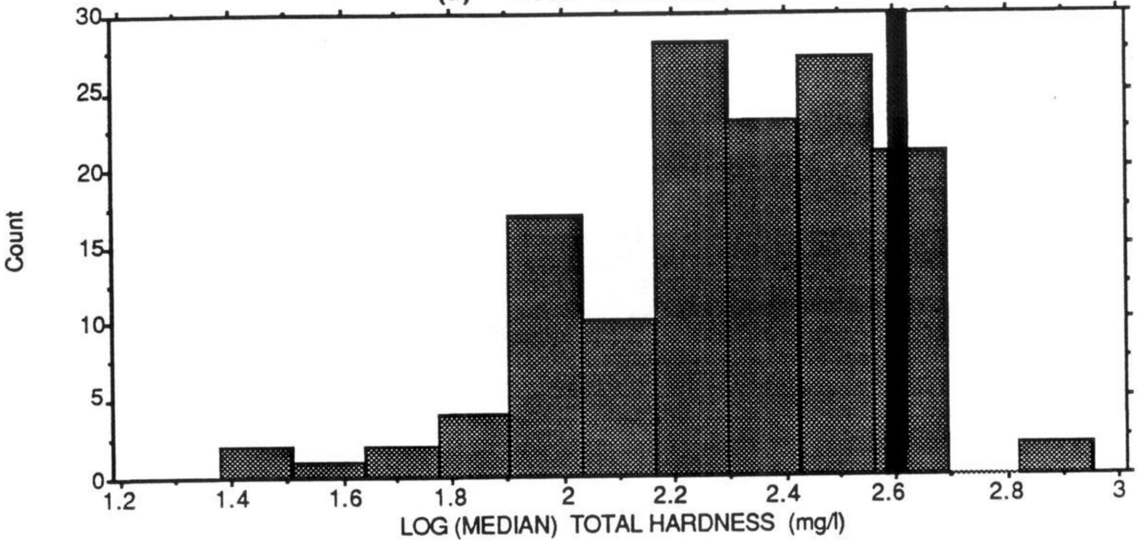
Interpretation of the results requires an understanding of the limitations of the analysis. The most important limitation is the fact that only indicators of change from background water quality were used; specific hazardous substances were not considered. Results should not be interpreted as an absolute ranking of water quality at landfills.

While this technique does have limitations, it can be used to prioritize sites for the setting of PAL's. The method considers both the degree and extent of contamination at the facilities. Degree is considered in the predictor itself, although only as an absolute assessment of "clean" or "dirty." Extent is addressed implicitly by the grouping criteria.

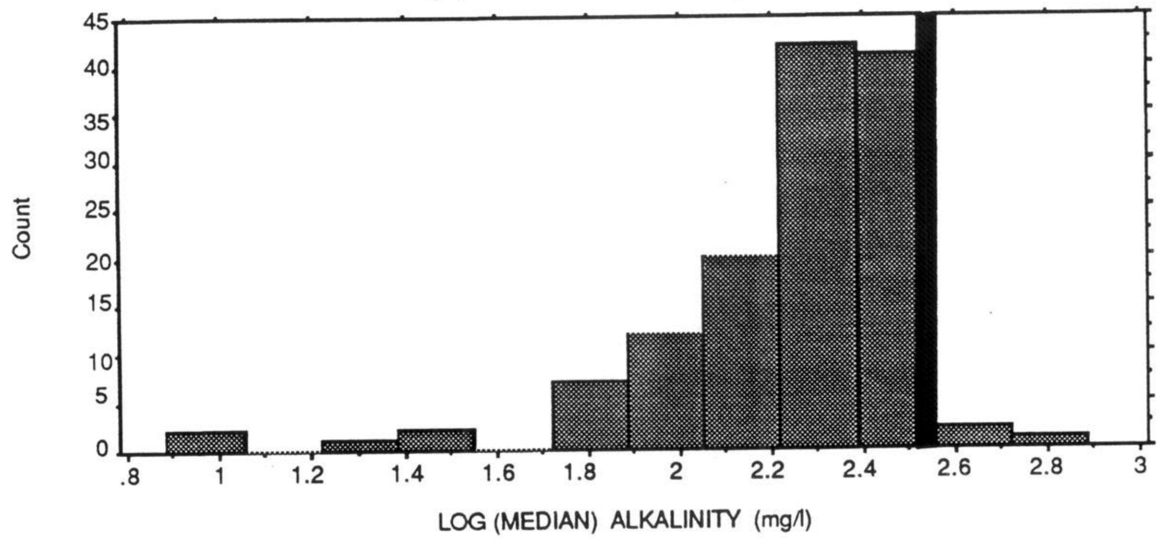
A second result of the exploratory data analysis is a description of characteristics of background water quality. The log-distributions of the median and fourth-spread may be used to get a preliminary indication of what typical "high" and "low" background concentrations are. These figures may be particularly useful to new personnel trying to interpret relative concentration levels from laboratory reports or time versus concentration plots.

Figure 4-7 Background Histograms for the Median

(a) Total Hardness



(b) Total Alkalinity



(c) Specific Conductance

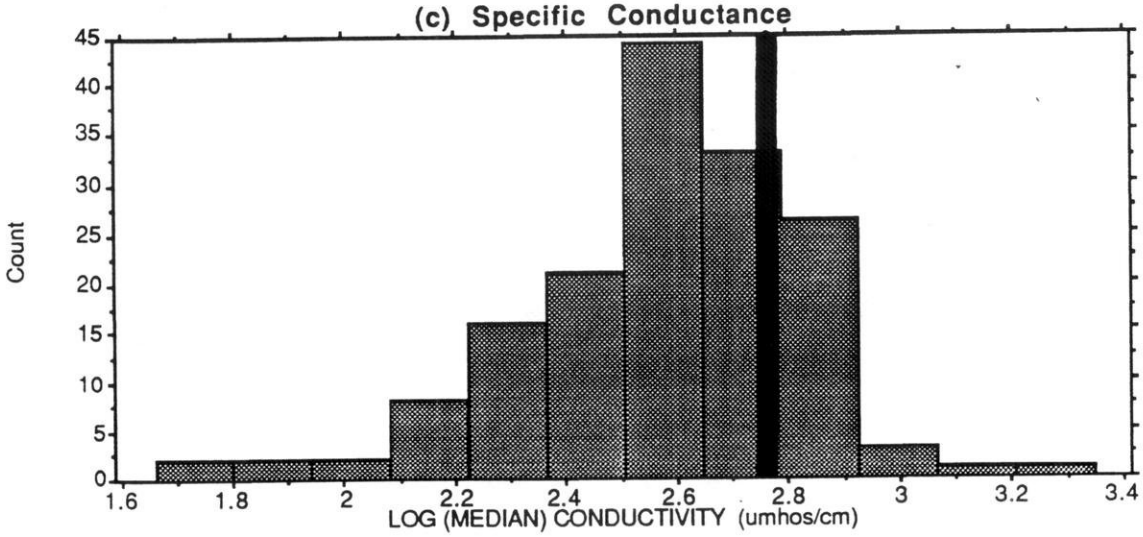
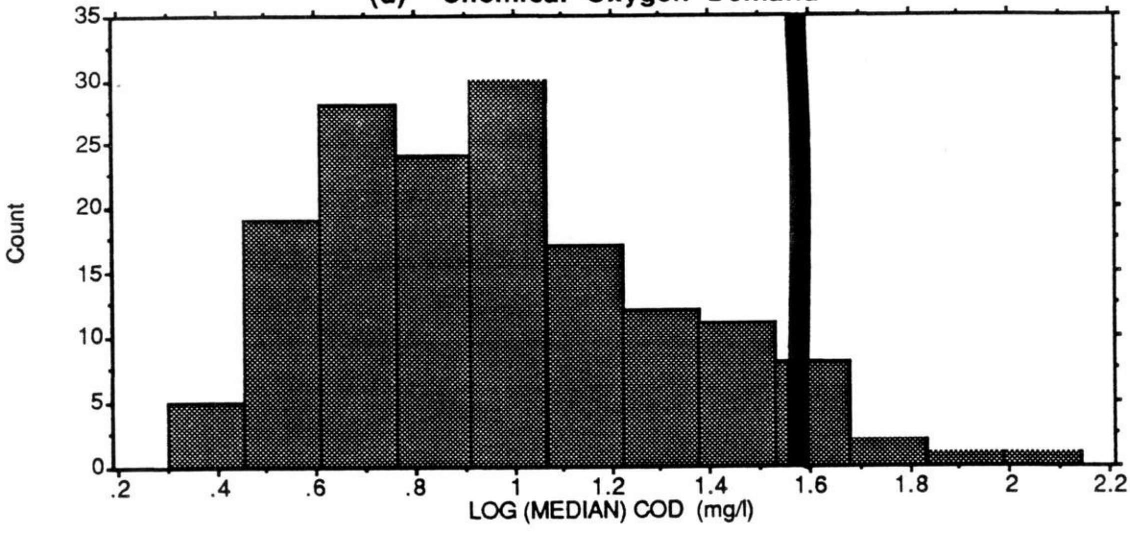
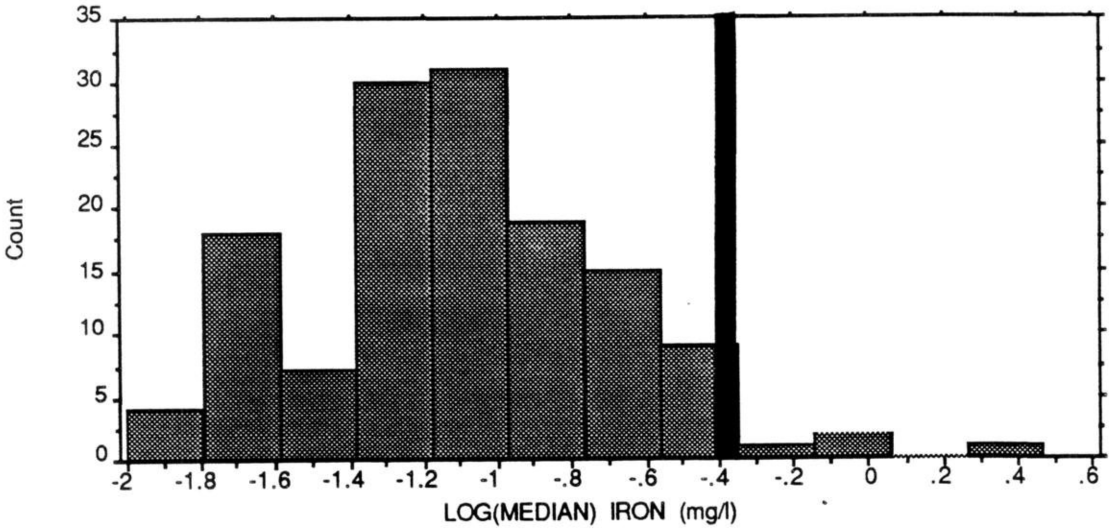


Figure 4-7 (continued)

(d) Chemical Oxygen Demand



(e) Iron



(f) Chloride

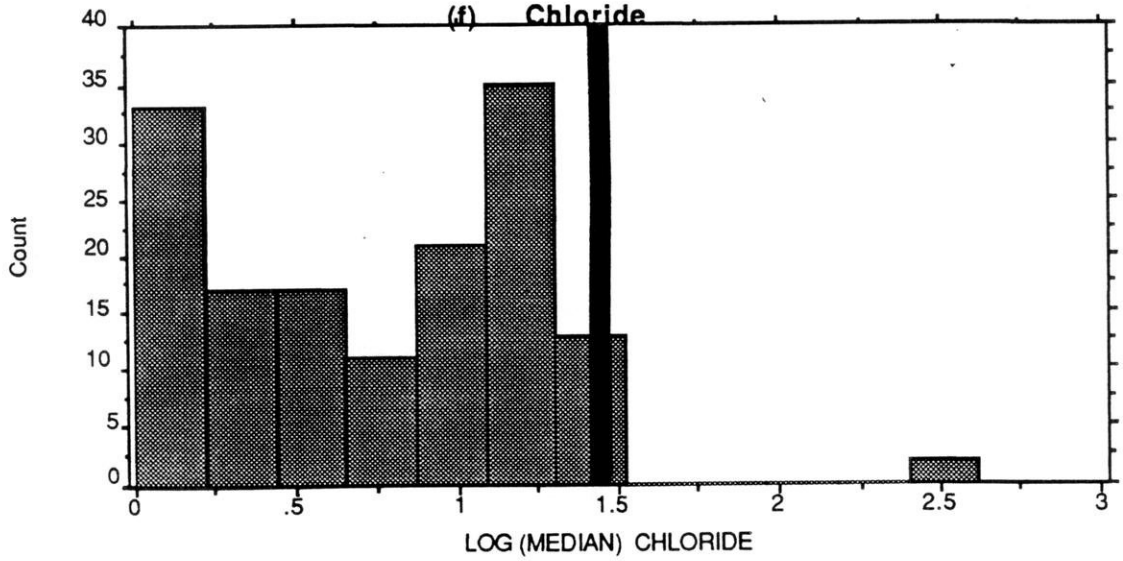


FIGURE 4-8 Background Histograms for the Interquartile Range

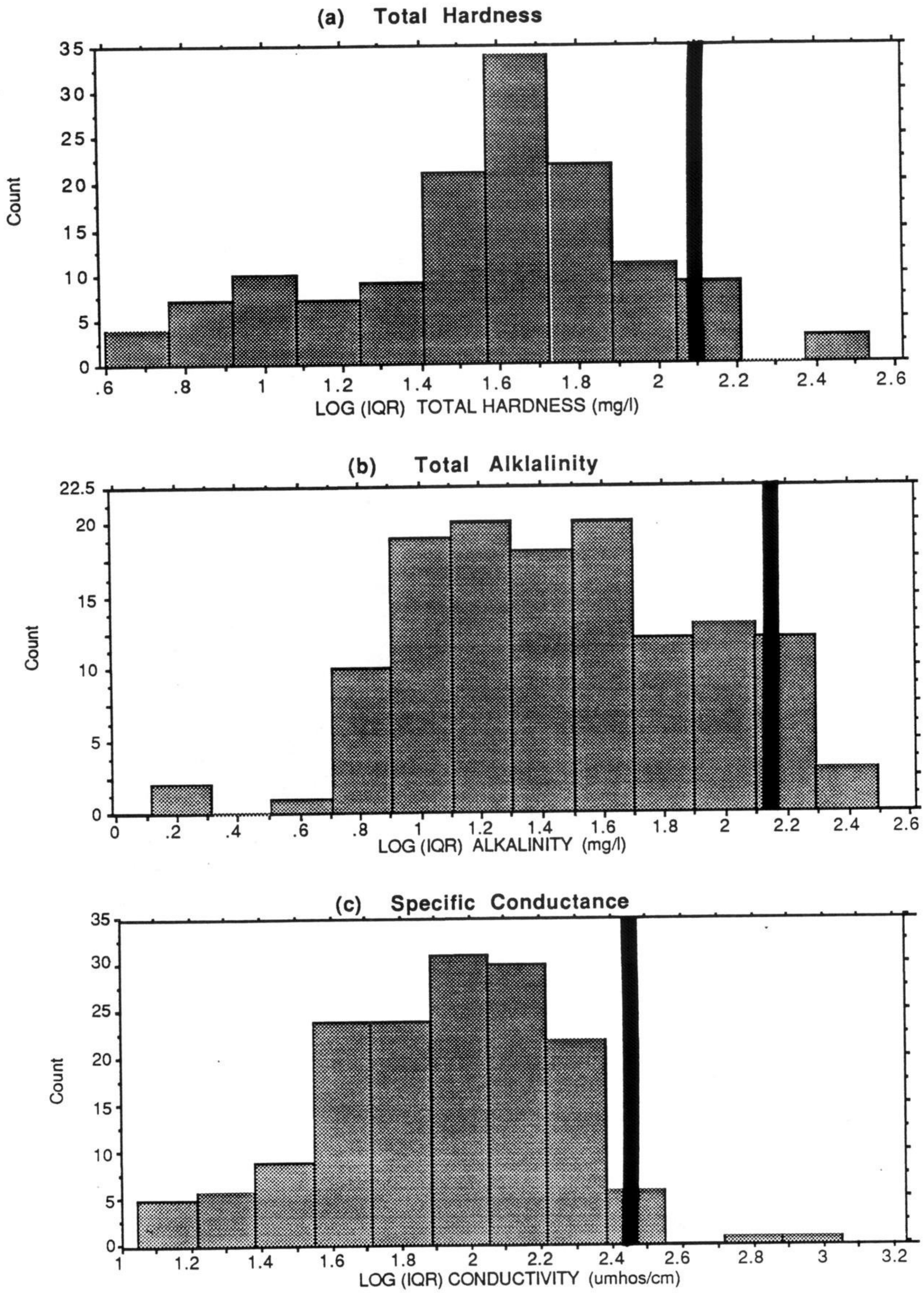
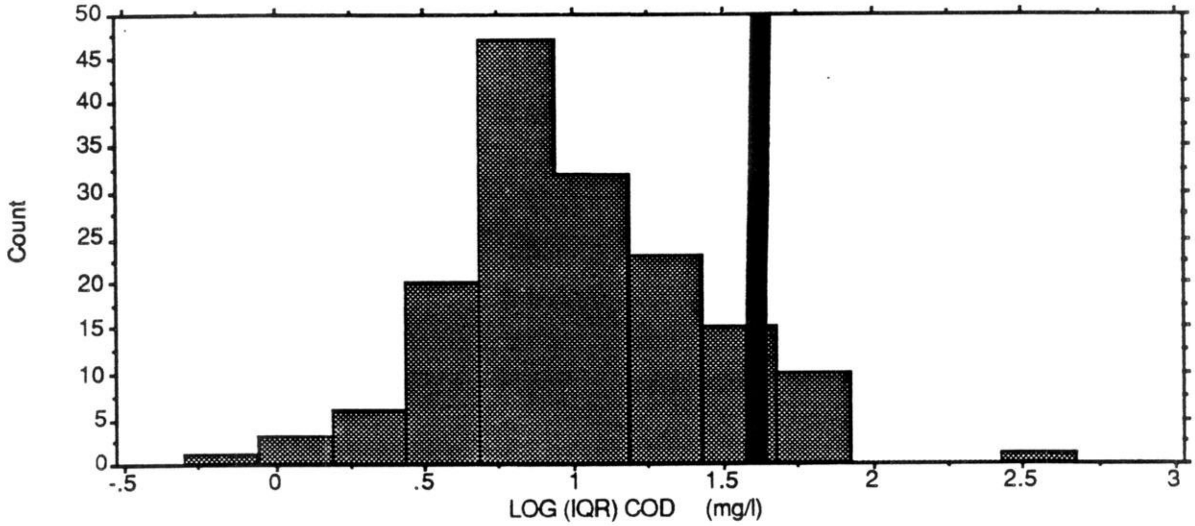
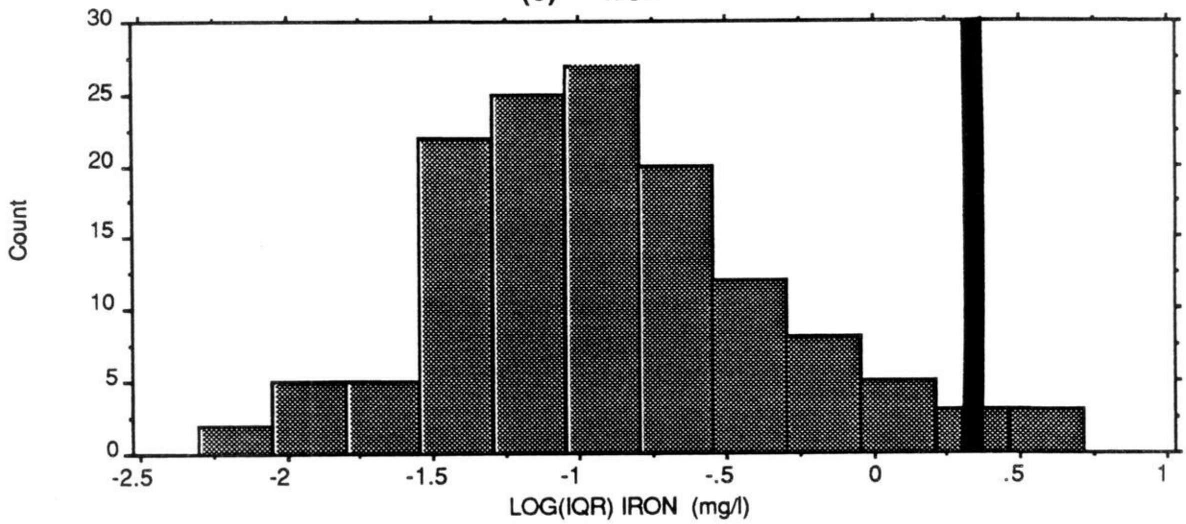


Figure 4-8 (continued)

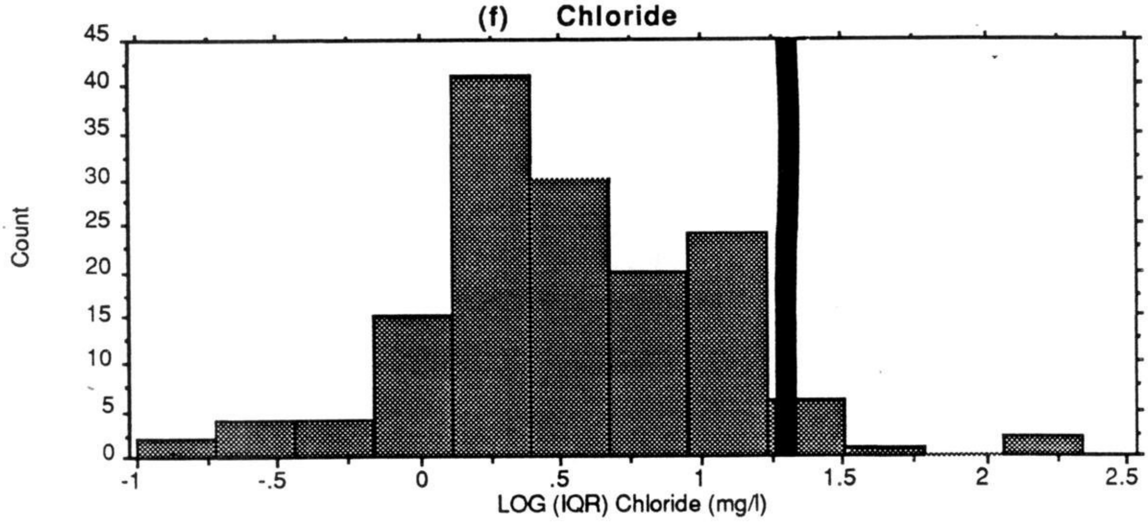
(d) Chemical Oxygen Demand



(e) Iron



(f) Chloride



CHAPTER FIVE
CONCLUSIONS AND RECOMMENDATIONS

5.0 Overview

This report has explored the use of graphics and statistics to detect groundwater contamination at waste disposal sites. Hydrogeology and water quality at twenty Wisconsin waste disposal sites were considered for this research. The sites represent a good cross-section of landfill design, size and geologic location for the State of Wisconsin.

This chapter summarizes the major findings of the previous chapters from a regulatory and technical perspective. Also, defensible procedures for evaluating contamination are synthesized into a general methodology useful in two general situations:

- existing waste sites with historic evidence of contamination, and
- new or existing sites with no suspected impact on groundwater quality.

The broad objective of this study is to evaluate alternative analytic methods that DNR can use to meet the intent of NR 140 at nonhazardous waste sites. To meet this objective, graphical and statistical methods appropriate for individual site reviews were evaluated in Chapters 2 and 3. Also, in order to help DNR prioritize their work in enforcing NR 140, a statistical screening of all licensed Wisconsin facilities is presented in Chapter 4. While the focus of this study is on solid waste disposal facilities, the insight provided is in many cases applicable to most types of hazardous waste sites, land disposal systems and storage facilities.

This chapter is organized as follows:

- the technical findings are discussed with respect to regulatory objectives in Section 5.1;
- our concerns with current EPA policy and technical guidance on statistical analysis of groundwater quality data are summarized in Section 5.2; and

- procedures to document groundwater contamination are synthesized into a general method in Section 5.3.

5.1 Summary and Regulatory Perspective

The most powerful analytic tools to detect groundwater contamination are graphs of water quality data. Well trained analysts can usually build a conclusive case for remedial action or further investigation based solely on geology, groundwater flow data, waste disposal history, and water quality graphs. Time versus concentration plots and boxplots are two powerful methods for detecting contamination. Time versus concentration plots are easy to prepare and may show trends in time and any abrupt changes in water quality. For these reasons, we recommend that DNR amend current solid waste and waste water regulations to require submittal of annual time series plots. Review of these submittals would be a quicker and more effective method of detecting new problems than trying to review all the laboratory analysis turn-around documents.

Because groundwater quality varies naturally both in space and in time, statistical methods are applicable to the regulatory decision making process. The correct use of statistical tests requires that assumptions implicit to the chosen statistical model be valid. Our investigation of the validity of assumptions for parametric statistical tests revealed that the assumption of normality may be frequently violated. Many of the most powerful statistical tests rely on this assumption. This finding further emphasizes the importance of graphical techniques and investigation of hydrogeologic conditions. It also supports the use of nonparametric statistics, although these methods can be less powerful than parametric counterparts. On a more positive note, results of previous studies show that seasonality and serial correlation are not frequently found in groundwater quality data. Thus we may conclude that the basic assumptions of stationarity and independence are valid.

Distributional assumptions aside, the use of statistical methods is complicated by spatial variations in background water quality. An analysis of 32 wells at three "clean" solid waste facilities showed that wells screened in similar geologic strata cannot be assumed to have equal mean or constant variance. When this is the case, groundwater contamination cannot be statistically discerned from natural variability (when comparisons are made between background wells and compliance wells). Whenever possible, background water quality should be defined for each well at a site.

The DNR relies on PAL's as indicators of potential groundwater contamination. At new and existing "clean" sites our research supports the DNR policy of setting PAL's at each well, rather than establishing site-wide levels. DNR also currently relies on PAL's as a tool to enforce regulations at existing sites with known contamination. In this case, PAL's are set at contaminated wells based on data from another well(s) at the site. for several reasons we recommend that DNR discontinue this practice. In the first place, at a clearly contaminated site, a case for remedial action can be built using water quality graphs, hydrogeologic information and good judgement. Calculation of PAL's is always secondary to this review. Secondly, transposing PAL's from one well or group of wells to another is not a sound statistical practice, since groundwater quality may significantly vary in space.

PAL's may be thought of as an upper limit for background water quality. Two other methods which are also used to define background water quality are statistical prediction intervals and tolerance intervals. A review of the statistical concept of these statistics shows that the prediction interval most directly compares background water quality to new data. The prediction interval is a background concentration estimated to have a set (low) probability of exceedance for new samples. The limit takes into account small sample size (unlike the PAL) and natural variation. A comparison of prediction intervals, tolerance intervals, and PAL's showed that the PAL is a conservative (high) estimate of background water quality when a significance level of 0.95 is employed in the calculations as recommended by EPA. It is important to note that these methods may not produce environmentally meaningful results if parametric distributional assumptions are not met. Since we know this is often the case, and because PAL's appear to be conservative, we recommend DNR continue to calculate background water quality levels based on the algorithm $PAL = \bar{x} + 3s$.

At hazardous waste sites (permitted under RCRA Subtitle C), EPA requires that quarterly data be compared to background water quality using a statistical test. Our investigation has shown that statistical prediction intervals are the preferred method. However as mentioned above, these intervals are dependent on the validity of distributional assumptions. Also, the interval should be updated annually as long as no contamination is present. We also recommend this method for sanitary landfills if (when) EPA adopts the amendments to Subtitle D of RCRA as proposed in 1988. The PAL does not appear to meet the requirements of the proposed regulations. The prediction interval calculation would replace comparison of new data to PAL's.

New monitoring data must be compared to mandated water quality standards as well as to background water quality. Wisconsin enforcement standards and PAL's set as a percent of these standards are interpreted in this report as being not-to-be-exceeded limits, except if sample collection, handling or laboratory error can be proven. The possibility of an exceedance being caused by sampling error or laboratory mishandling can be assessed by the timely resampling of the well. Other than that, the only reason that a sample value above the standard may not be a true exceedance is if laboratory accuracy is a factor. We suggest that sample values close to a standard be given the benefit of the doubt only if a lower accuracy confidence range reported by the laboratory indicates that the true sample concentration is below the standard. DNR should further investigate the use of laboratory accuracy reports for determining standard exceedances.

The above discussion focused on the main issues surrounding individual site review. The graphical and statistical procedures which we recommend are synthesized into a general methodology in Section 5.3.

Our research also resulted in the development of a predictor of groundwater quality change. A control group of 161 background wells and 113 "contaminated" wells was used to evaluate the performance of the predictor. This predictor separates "clean" and "dirty" wells based on concentration limits for both the median and interquartile range of a dataset below which contamination was unlikely. These nonparametric statistics were found to be consistently low for "clean" wells, unlike parametric distribution parameters which are sensitive to data outliers. Ongoing DNR efforts should be focused on those sites which appear to significantly impact groundwater and on sites which are believed to need additional monitoring. A secondary result of the predictor research was a characterization of clean water quality. Between the statistical screening results and the insight gained from the evaluation of statistical tests, the DNR hopefully is in a better position to prioritize and expedite the remediation of contaminated groundwater resources in Wisconsin.

5.2 Concerns with EPA Policy and Recommendations

The final EPA rules for permitted RCRA hazardous waste facilities (EPA, October 11, 1988) require a quarterly test for "change" in water quality as well as comparison of new data to water quality standards. A draft guidance document for these regulations is currently in the final review stage (EPA, 1988). This document is a statistical "cookbook" for hazardous waste facility owners. It is also intended to provide guidance for the proposed changes to RCRA Subtitle D which affect municipal solid waste facilities. Throughout this report we evaluated and discussed procedures recommended in this document. In this section our primary concerns with the guidance document are summarized. These concerns are:

- the lack of insight on environmental questions which recommended tests are "answering;"
- EPA policy of not requiring tests for distributional assumptions prior to use of parametric statistics;
- recommendation of parametric ANOVA as a "default" method of analysis; and
- use of confidence and tolerance intervals for detecting standard exceedances.

The document does not discuss the advantages and disadvantages of the various recommended procedures. Looking at the broad picture, the choice between making between-well comparisons versus intra-well tests is not thoroughly discussed. The document implies that between-well comparisons are preferred. Conversely, we feel that intra-well comparisons are statistically more correct than inter-well tests sincethe latter are confounded by natural spatial variability. From a closer perspective, the recommended tests for both inter- and intra-well comparisons should be presented by defining the null and alternative hypotheses in terms of the "question" the test answers as well as in statistical notation. Some insight should be provided regarding the fundamental difference in philosophy of the recommended procedures.

Addressing the second issue listed above, we believe that tests for normality should be performed prior to using parametric statistical tests. This research has shown that this assumption is frequently violated by groundwater quality data, even after log-transformation. Three (optional) tests for normality are detailed in the EPA document: normal probability plots, the coefficient of variation method, and the chi square test. We do not believe that any of these tests are appropriate for wide-spread use to test for normality, although the probability plots are useful for qualitative inferences. The chi square test is not very sensitive to departures from normality at small sample size. The coefficient of variation method is not a "goodness-of-fit" test, rather the procedure is a "rule of thumb." We found this rule to grossly underestimate the number of non-normal datasets found by the skewness test. In place of these methods we recommend that EPA advocate the skewness test. This test is simple to perform and is sensitive to small sample size.

Our most serious concern with the EPA guidance document concerns the use of between-well comparisons, and particularly the ANOVA method. In this study, 8 parameters measured at 4 groups of wells (from three sites) were tested for natural spatial variability. The assumptions implicit in parametric ANOVA were tested for both the raw data and the log-transformed data. This analysis found that

- the assumption of constant variance (as tested by Bartlett's test) was found to be frequently violated for both the raw and log-transformed data; and

- the normality assumption was also frequently violated.

Based on the above findings, the Kruskal-Wallis test (nonparametric one-way ANOVA) was used to test whether mean water quality differed between background wells. At all four groups of wells spatial variability was detected for at least two parameters. From these results we conclude that shifts in mean due to contamination cannot be distinquished from natural shifts in mean water quality at these sites. Therefore, we believe that neither parametric nor nonparametric ANOVA should be used to test for contamination, unless it can be shown that natural spatial variability is not significant.

Another important issue regards the use of confidence and tolerance intervals for detecting standard exceedances. First of all, the use of an *upper* confidence limit (on the mean of four independent quarterly samples) is a lot less protective of the environment than the alternative method of using a *lower* tolerance interval. The confidence limit method could allow three out of four samples to be above a standard and still not conclude that the standard has been exceeded. This is because very wide intervals are calculated when sample variance is large, which is usually the case for "contaminated" samples. Secondly, both of these methods rely on calculating the mean and standard deviation from samples which may be contaminated by landfill leachate. We do not believe this is correct because samples drawn from a contaminated regime are not necessarily from the same population, since leachate plumes are not homogeneous mixtures evenly dispersed in groundwater. Thirdly, as discussed in Section 3.3, <u>we are not sure that statistical intervals should play a role in determining standard exceedance at all, except for perhaps considering laboratory accuracy confidence levels.</u>

More detailed discussion of EPA recommended tests and policy may be found throughout this report and in particular in,

- Section 1.3 Federal Regulatory Context,
- Section 2.2.3 The Assumption of Normality,
- Section 3.1 Tests of Central Tendency, and
- Section 3.3 Confidence, Tolerance and Prediction Intervals.

5.3 Statistical Procedures to Document Groundwater Contamination

The results of this research indicate that the following procedures are preferred for evaluating contamination at landfills with respect to determining compliance with groundwater quality regulations:

- time versus concentration plots and box plots,

- tests of trend to support cases of contamination,
- resampling of wells,
- comparison of standard exceedances to laboratory accuracy confidence levels, and
- PAL's and prediction limits for detecting high concentrations thought to be indicative of contamination.

When parametric statistics are employed (for example for prediction intervals) we recommend,

- the skewness test for testing normality, and
- log transformation of data if raw data are positively skewed.

These methods are synthesized into a methodology for analyzing water quality at waste disposal sites on Figures 5-1 to 5-5. No single method is appropriate at all sites. The guidance provided in this section is intended to have general applicability for determining if background water quality has significantly changed at waste disposal facilities. Basically the approach is to define background water quality using hydrogeologic information, summary statistics and review of water quality graphs. If a site appears clean or is a new site, a future comparison procedure is presented. At existing sites methods are shown for documenting existing contamination if present.

Figure 5-1 summarizes the overall approach. Procedures to evaluate background water quality at existing sites are presented in Figure 5-2. Time versus concentration plots and boxplots are used to identify those wells which are believed to be clean. The DNR currently has adopted an approach recommended by Goodman and Potter (1987) for standardizing well data in order to plot all parameters at a well on one plot. This method uses nonparametric statistics based on the median and IQR to transform data from different parameters to a single (NP) scale. This procedure is termed "optional" on Figure 5-2 since it is not as essential as the one- parameter time series plot or the box plot. Also, the NP scale is not as easy to interpret as a true concentration scale.

We suggest that time series plots be annotated by noting the presence of siginificant trends and that box plots be annotated by adding the sample size and the number of standard exceedances. If plots are rescaled to "hide" high data this should also be noted on the figure. We do not recommend the arbitrary deletion of high data values just because outliers are infrequent. Unless editing of data can be physically justified, it should not be done.

Statistics should be summarized for each well. Important summary statistics are listed on Figure 5-3. In addition to familiar distributional parameters we recommend that Kendall's Tau statistic and the skewness coefficient be generated on a routine basis. Kendall's Tau is a measure of temporal correlation and, if found to be significant, indicates that a trend exists. The procedure to apply Kendall's test of trend is described in Appendix A. Also described in Appendix A is the skewness test.

The statistical program used by the DNR to generate statistics is the STATISTICAL ANALYSIS SYSTEM (SAS). Currently DNR has programed SAS to generate output for a multitude of summary statistics for each well. Most of these statistics such as the coefficient of variation, the quantile points, etc are not reviewed. We suggest that DNR reprogram SAS to output only those statistics listed in Figure 5-3. This new output format will be easier to interpret and probably inspected in more detail than the current forms.

Figure 5-4 presents procedures for documenting existing contamination at a site. For sites with apparent contamination we see no need to calculate PAL's. The only statistical test we suggest is Kendall's test for trend. Significant positive trend coupled with standard exceedances is powerful evidence of contamination. For sites where more extensive documentation of contamination is necessary, we suggest that PAL's be calculated as currently mandated. A viable alternative however is the statistical prediction interval. The procedure for calculating a prediction interval is given in Appendix A. Calculation of a prediction interval would also be appropriate at sites when a quarterly comparison to background water quality is required. The skewness test for normality should be applied before calculating prediction intervals. It is important to recognize that PAL's and prediction intervals calculated on non-normal data sets may not be environmentally meaningful. Log-transformation of data should be considered in this situation.

A method to evaluate quarterly reports submitted by supposedly "clean" sites is presented in Figure 5-5. The procedure advocates resampling of wells if sample or gross laboratory error are suspected. A comparison of health and welfare parameters to laboratory accuracy confidence levels is recommended.

These admittedly general procedures are presented so that site-review personnel will have a set of tools which provide the level of scientific detail necessary to defensibly document groundwater contamination. We suggest that these procedures are defensible because 1) they are based primarily on graphical inference and good judgement and 2) nonparametric statistics are advocated whenever possible (i.e. trend tests and box plots). The one parametric method suggested for general use is the prediction interval. Because data may not be normally or lognormally distributed, a test for normality is recommended.

These flow charts coupled with the explanation of graphical procedures in Chapter 2, the discussion of statistical tests in Chapter 3, and the statistical methods described in Appendix A provide DNR with a foundation for determining compliance with NR 140 at solid waste disposal facilties. The sites targeted by the statistical screening of the Wisconsin groundwater quality database should be given high priority for investigation using these procedures.

FIGURE 5-1 Overview of Methodology

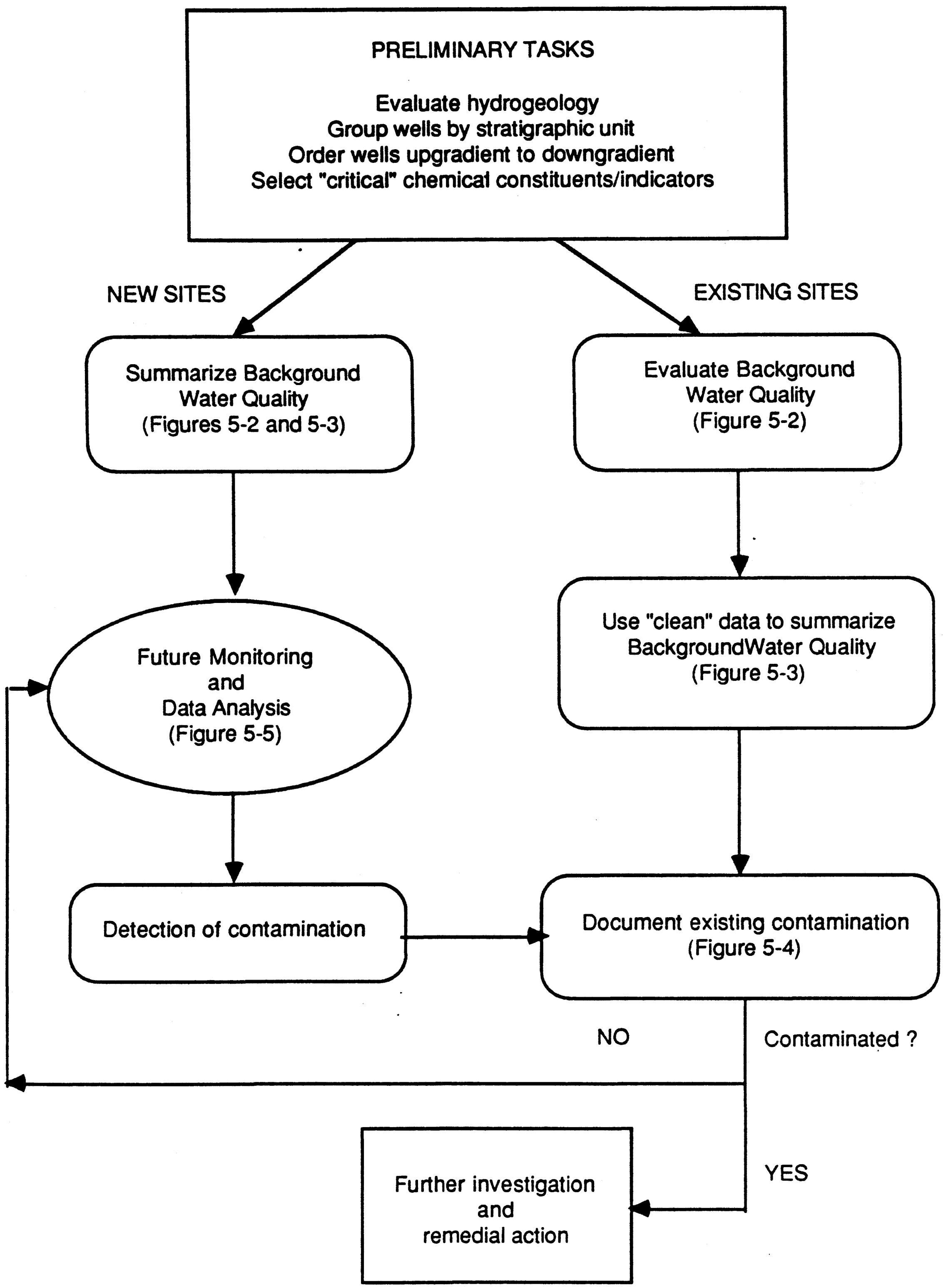


FIGURE 5-2 Evaluation of Background Water Quality at Existing Sites

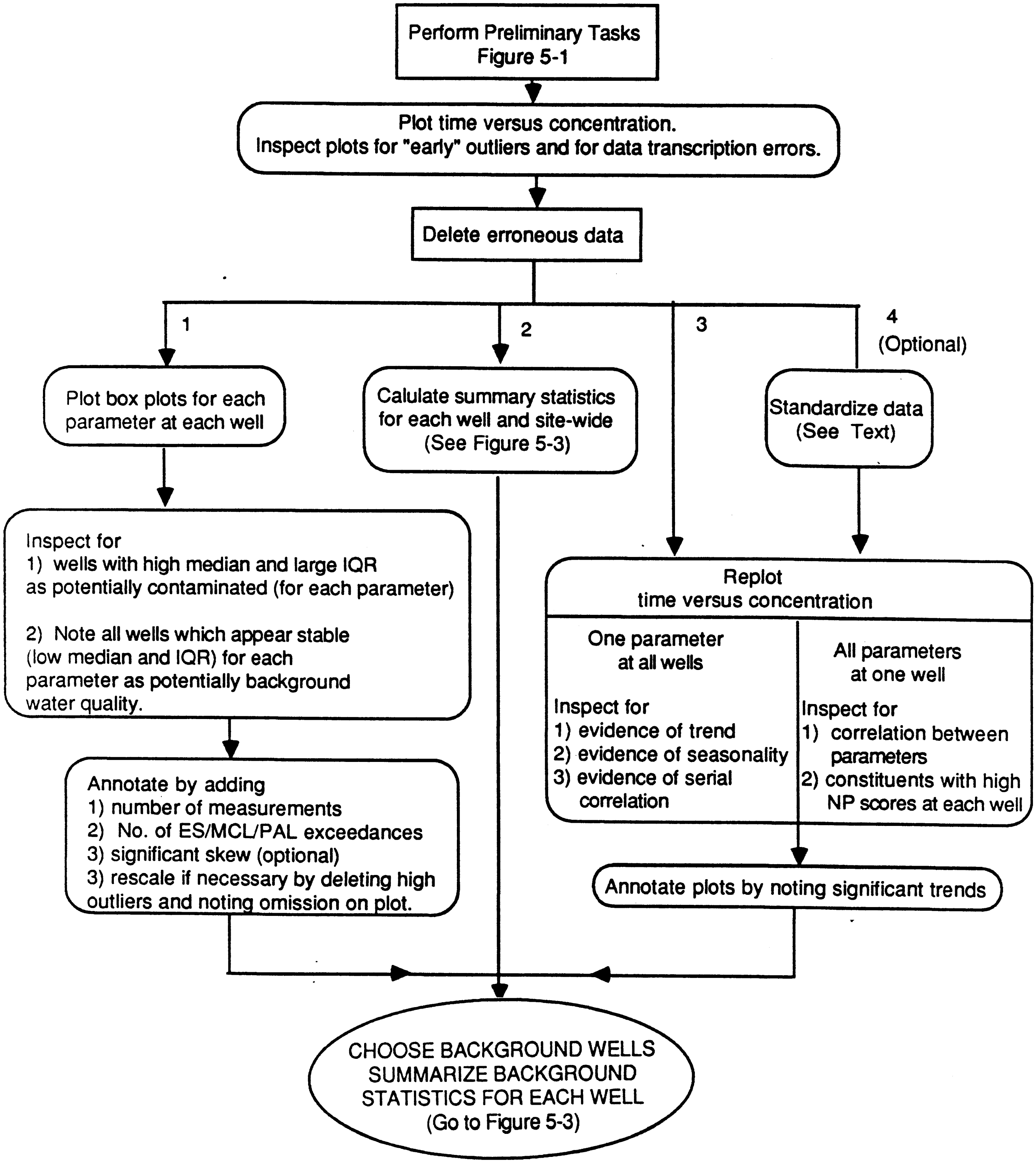


FIGURE 5-3 SUMMARY OF BACKGROUND WATER QUALITY
(Existing sites and new sites)

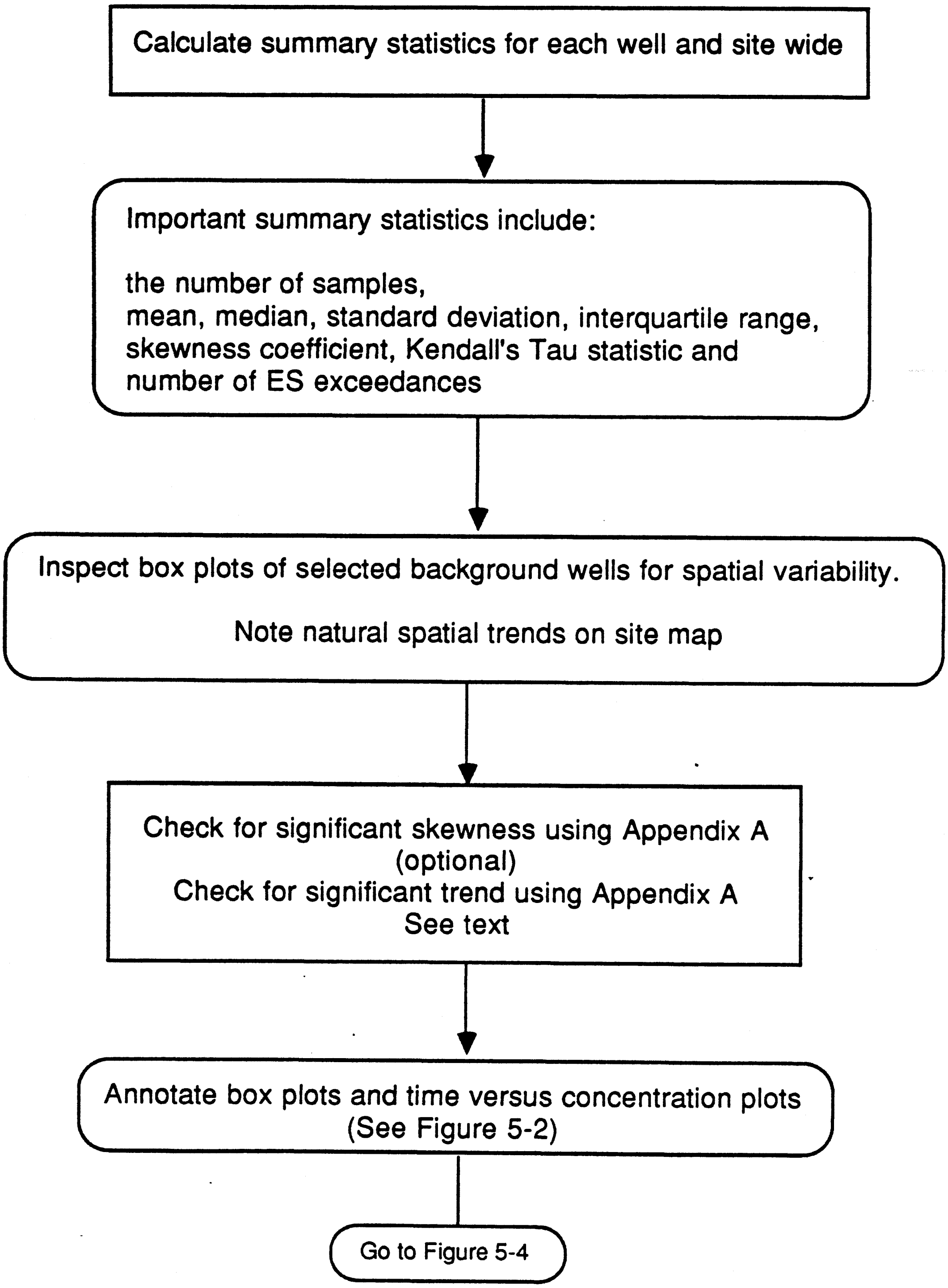


Figure 5-4 Document Existing Contamination and/or set Limits

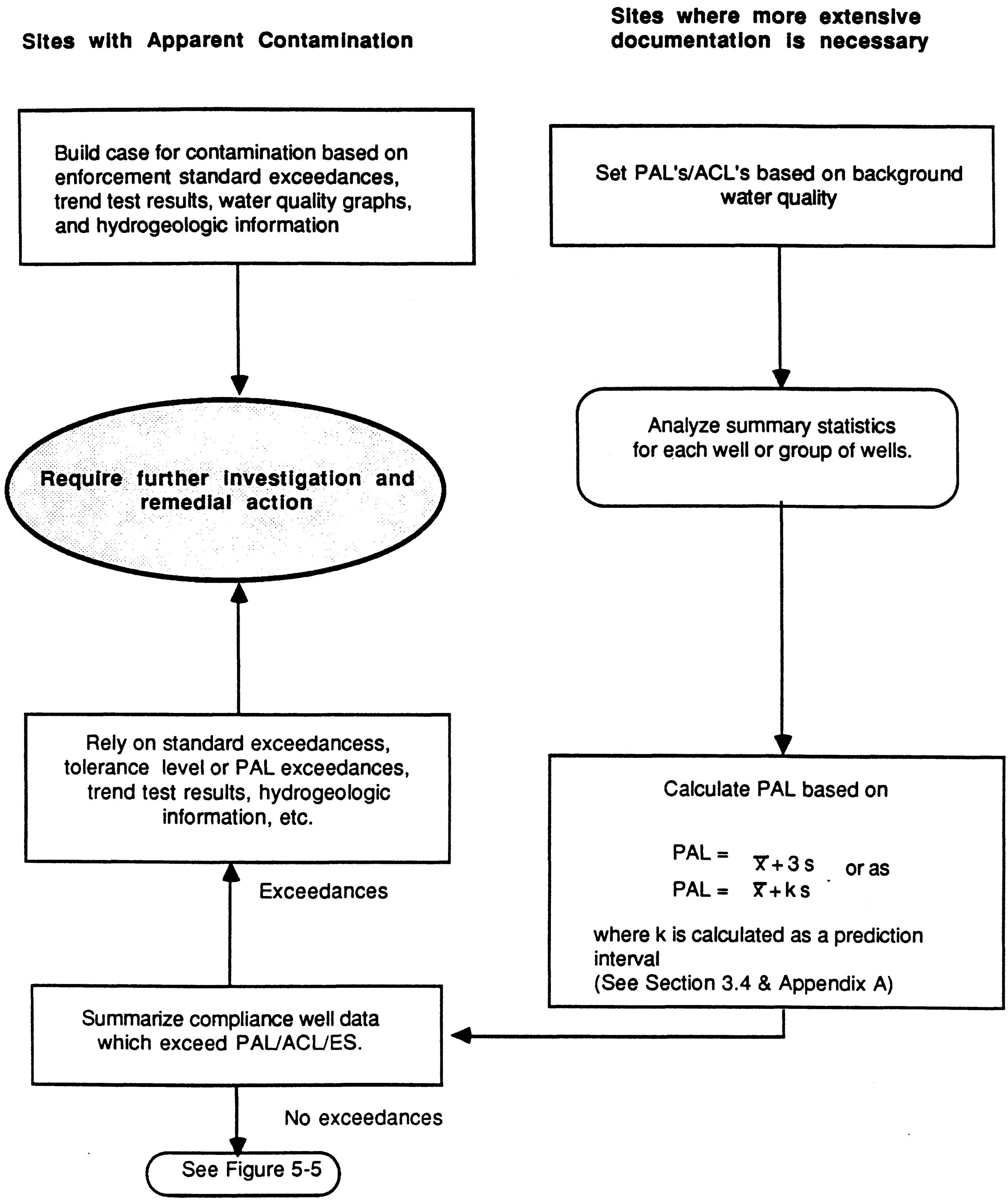
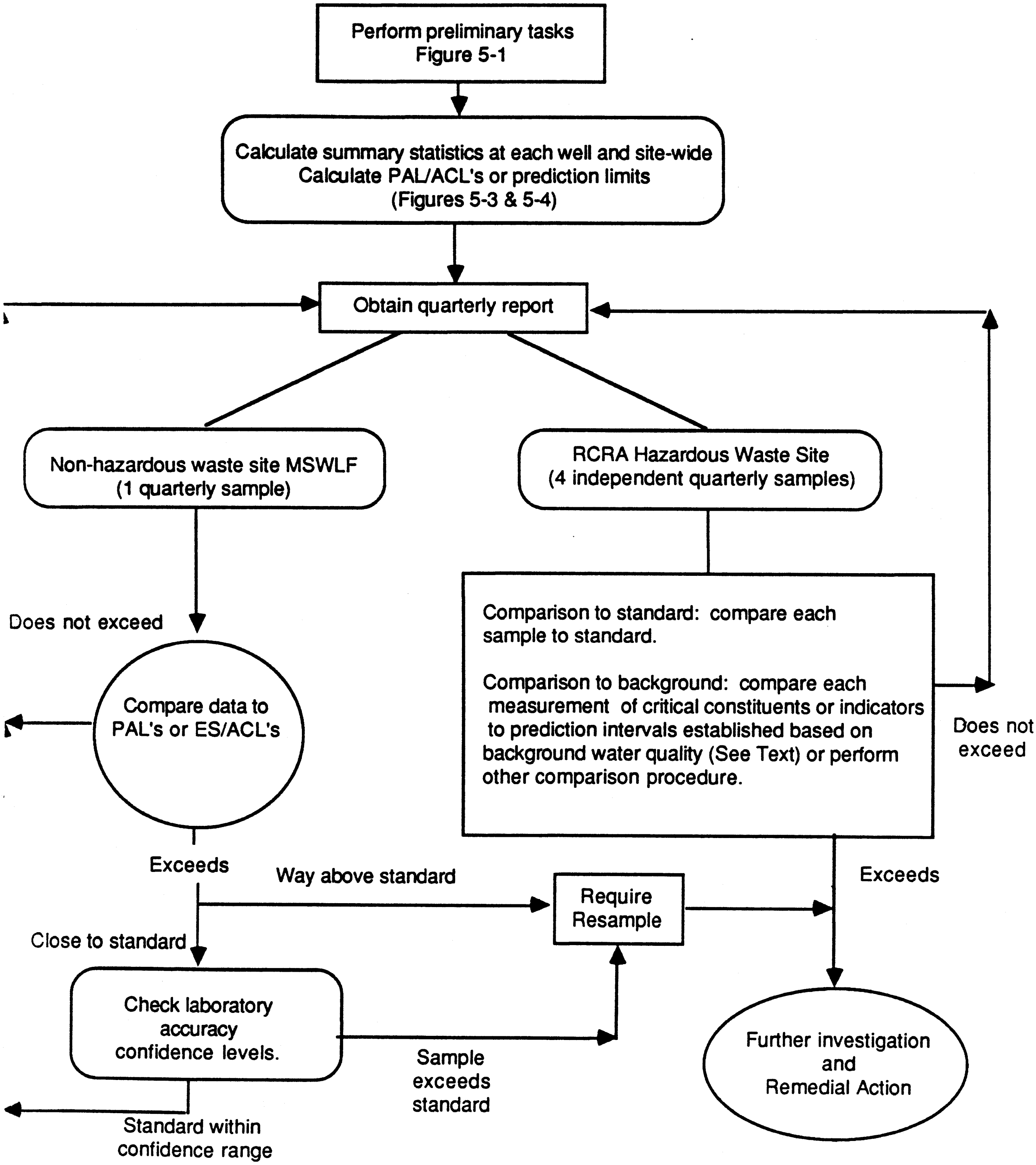


Figure 5-5 Determining Compliance at "Clean" Sites



REFERENCES

Benjamin J.R. and C. A. Cornell. 1970. Probability, Statistics, and Decision for Civil Engineers. New York: McGraw-Hill Book Company

Berthouex, P.M. and W.G. Hunter. 1983. How to Construct Reference Distributions to Evaluate Treatment Plant Effluent Quality. Journal WPCF. Vol. 55, No. 12 (December): 1417-1423.

Blair, R. Clifford and James J. Higgins. 1980. A Comparison of the Power of Wilcoxon's Rank-Sum Statistic to that of Student's T Statistic Under Various Nonnormal Distributions. Journal of Educational Statistics. Vol 5, No. 4 (Winter): 309-335.

Boneau, C. Alan. 1960. The Effects of Violations of Assumptions Underlying the T-Test. Psychological Bulletin. Vol. 37, No. 1: 49-64.

Box, G.E.P., W.G. Hunter and J.S. Hunter. 1978. Statistics for Experimenters. New York: John Wiley & Sons.

Cohen, A.C., Jr. 1959. Simplified Estimators for the Normal Distribution When Samples are Singly Censored or Truncated. Technometrics. Vol 1: 217-237.

Cohen, A.C., Jr. 1961. Tables for Maximum Likelihood Estimates: Singly Truncated and Singly Censored Samples. Technometrics. Vol. 3: 535-541.

Conover, W.J. and R.L. Iman. 1976. On Some Alternative Procedures Using Ranks for the Analysis of Experimental Designs. Commun. Statist.-Theor. Meth., Vol. A5, No. 14: 1349-1368.

Davis, C.B. and R.J. McNichols. Discussion of Statistical Prediction Intervals for the Evaluation of Ground-Water Quality, by R.D. Gibbons. Ground Water. : 90-91.

Doctor, P.G., R.O. Gilbert, R.A. Saar, and G. Duffield. 1985a (February 14). An Analysis of Sources of Variation in Ground-Water Monitoring Data of Hazardous Waste Sites. Milestone 1. Revised Draft. EPA Contract No. 68-01-6871. Battelle, Pacific Northwest Laboratories. Richland, WA.

Doctor, P.G., R.O. Gilbert, and R.R. Kinnison, 1985b (February 25). Ranges of Variation in Ground-Water at Hazardous Waste Sites. Draft Milestone 2. EPA Contract No. 68-01-6871. Battelle, Pacific Northwest Laboratory. Richland, WA.

Doctor, P.G., R.O. Gilbert and R.R. Kinnison 1985c (May 9). Statistical Comparisons of Ground-Water Monitoring Data-Draft. Milestone 3. EPA Contract No. 68-01-6871. Batelle, Pacific Northwest Laboratory. Richland, WA.

EPA (United States Environmental Protection Agency). August 24, 1987. Statistical Methods for Evaluating Ground-Water Monitoring Data from Hazardous Waste Facilities; Proposed Rule. 40 CFR Part 264. Federal Register (Part VII). Vol. 52, No. 163 (Monday).

EPA (United States Environmental Protection Agency). October 11, 1988. Statistical Methods for Evaluating Ground-Water Monitoring from Hazardous Waste Facilities; Final Rule. 40 CFR Part 264. Federal Register (Part II). Vol. 53, No. 196 (Tuesday).

EPA (United States Environmental Protection Agency). 1988. Statistical Analysis of Ground Water at RCRA Facilities-Draft Guidance. Office of Solid Waste. Waste Management Division. Washington, D.C.

Fligner, Michael A. and George E. Policello II. 1981. Robust Rank Procedures for the Behrens-Fisher Problem. Journal of the American Statistical Association. Vol. 76, No. 373. (March): 162-168.

Florida, Department of Environmental Regulation. June, 1985. Statistical Procedures and Considerations for Environmental Management (SPACEMAN). DER Technical Report by L. Groeneveld and R.D. Duval.

Fung, Karen Yuen. 1979. A Monte Carlo Study of the Studentized Wilcoxen Statistic for the Behrens-Fisher Problem. Journal Statist. Comput. Simul. Vol. 10: 15-24.

Gibbons, Jean D. 1985a. Nonparametric Methods for Quantitative Analysis. Columbus, Ohio: American Sciences Press, Inc.

Gibbons, Jean D. 1985b. Nonparametric Statistical Inference. New York: Marcel Dekker, Inc.

Gibbons, Robert D. 1987a. Statistical Prediction Intervals for the Evaluation of Ground-Water Quality. Ground Water. Vol. 25, No. 4 (July-August): 455-465.

Gibbons, Robert D. 1987b. Statistical Models for the Analysis of Volatile Organic Compounds in Waste Disposal Sites. Ground Water. Vol. 25, No. 5 (September-October): 572-580.

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

Goodman, Iris. 1987. Graphical and Statistical Methods to Assess the Effect of Landfills on Groundwater Quality. M.S. Thesis. University of Wisconsin-Madison.

Goodman, Iris and Kenneth Potter. 1987. Graphical and Statistical Methods to Assess the Effects of Landfills on Groundwater Quality. Report to Wisconsin Department of Natural Resources, Bureau of Solid and Hazardous Waste (see also Goodman, 1987).

Hahn, Gerald J. 1970a. Statistical Intervals for a Normal Population, Part I. Examples & Applications. Journal of Quality Technology. Vol. 2, No. 3 (July): 115-125.

Hahn, Gerald J. 1970b. Statistical Intervals for a Normal Population, Part II. Formulas, Assumptions, Some Derivations. Journal of Quality Technology. Vol. 2, No. 4 (October): 195-206.

Hahn, Gerald. 1970c. Understanding Statistical Intervals. Industrial Engineering. Vol. 2, No. 12 (December): 45-48.

Hahn, Gerald J. and Wayne Nelson. 1973. A Survey of Prediction Intervals and Their Applications. Journal of Quality Technology. Vol. 15, No. 4 (October): 178-188.

Harris, Jane, Jim C. Loftis and Robert H. Montgomery. 1987. Statistical Methods for Characterizing Ground-Water Quality. Ground Water. Vol. 25, No. 2 (March-April): 185-193.

Helsel, Dennis R. 1987. Advantages of Nonparametric Procedures for Analysis of Water Quality Data. Hydrological Sciences Journal. Vol. 32, No. 2 (June): 179-190.

Helsel, Dennis R. and Robert M. Hirsch. 1988. Discussion of Applicability of the T-Test for Detecting Trends in Water Quality Variables by R.H. Montgomery and J.C. Loftis. Water Resources Bulletin. Vol. 24, No. 1: 201-207.

Helsel, Dennis R. and Timothy A. Cohn. 1988. Estimation of Descriptive Statistics for Multiply Censored Water Quality Data. Water Resources Research. Vol.24, No. 12. (December): 1997-2004.

Hirsch, Robert M. and James R. Slack. 1984. A Nonparametric Trend Test for Seasonal Data with Serial Dependence. Water Resources Research. Vol. 26, No. 6 (June): 727-732.

Hirsch, Robert M., James R. Slack and Richard A. Smith. 1982. Techniques of Trend Analysis for Monthly Water Quality Data. Water Resources Research. Vol. 18, No. 1 (February): 107-121.

Lettenmaier, Dennis P. 1976. Detection of Trends in Water Quality Data from Records with Dependent Observations. Water Resources Research. Vol. 12, No. 5 (October): 1037-1046.

Loftis, Jim C., Jane Harris and Robert H. Montgomery. 1987. Detecting Changes in Ground Water Quality at Regulated Facilities. Ground Water Monitoring Review. (Winter 1987): 72-76.

Loftis, J.C. and R.C. Ward. 1980. Sampling Frequency Selection for Regulatory Water Quality Monitoring. Water Resources Bulletin. Vol. 16, No. 3 (June): 501-507.

McBean, Edward A., Michel Kompter and Frank Rovers. 1988. A Critical Examination of Approximations Implicit in Cochran's Procedure. Ground Water Monitoring Review. Vol. ___ (Winter): 83-87.

McBean, Edward and Frank A. Rovers. 1984. Alternatives for Handling Detection Limit Data in Impact Assessments. Ground Water Monitoring Review. Vol. 4, No. 2 (Spring): 42-44.

McBean, Edward and Frank A. Rovers. 1984. Alternatives for Assessing Significance of Changes in Concentration Levels. Ground Water Monitoring Review. Vol. 4, No. 3 (Summer): 39-41.

McBean, Edward and Frank A. Rovers. 1985. Analysis of Variances as Determined from Replicate vs. Successive Sampling. Ground Water Monitoring Review. Vol. 5, No. 3 (Summer): 61-64.

McGill, Robert, John W. Tukey and Wayne A. Larsen. 1978. Variations of Box Plots. The American Statistician. Vol. 32, No. 1 (February): 12-16.

Millard, S.P. and S.J. Deverel. 1988. Nonparametric Statistical Methods for Comparing Two Sites Based on Data with Multiple Nondetect Limits. Water Resources Research. Vol.24, No. 12 (December): 2087-2098

Miller, M.D. and F.C. Kohout. (Undated.) RCRA Ground Water Monitoring Statistical Comparisons: A Better Version of Student's T-Test. Mobil Research and Development Corporation. Paulsboro, New Jersey.

Montgomery, Robert H. and Jim C. Loftis. 1987. Applicability of the T-Test for Detecting Trends in Water Quality Variables. Water Resources Bulletin. Vol. 23, No. 4: 653-662.

Montgomery, Robert H., Jim C. Loftis and Jane Harris. 1987. Statistical Characteristics of Ground-Water Quality Variables. Ground Water. Vol. 25, No. 2 (March-April): 176-184.

Montgomery, Robert H. and Kenneth H. Reckhow. 1984. Techniques for Detecting Trends in Lake Water Quality. Water Resources Bulletin. Vol. 20, No. 1 (February): 43-51.

NCASI. 1985. Groundwater Quality Data Analysis. National Council of the Paper Industry for Air and Stream Improvement, Inc. Technical Bulletin #462 (June).

Nelson, James D. and Robert C. Ward. 1981. Statistical Considerations and Sampling Techniques for Ground-Water Quality Monitoring. Ground Water. Vol. 19, No. 6 (November-December): 617-625.

Potthoff, Richard F. 1963. Use of the Wilcoxon Statistic for a Generalized Behrens-Fisher Problem. Annals of Mathematical Statistics. Vol. 34: 1596-1599.

Pratt, John W. 1964. Robustness of Some Procedures for the Two-Sample Location Problem. Journal of the American Statistical Association. Vol. 59 (September): 665-680.

Rice, George, James Brinkman and Dianna Muller. 1988. Reliability of Chemical Analyses of Water Samples-the Experience of the UMTRA Project. Ground Water Monitoring Review. (Summer 1988): 71-75.

Rovers, Frank A. and Edward A. McBean. 1981. Significance Testing for Impact Evaluation. Ground Water Monitoring Review. Vol. 1, No. 3 (Summer): 39-43.

Sen, P.K. 1968. Estimates of the Regression Coefficient Based on Kendall's Tau. Journal of the American Statistical Association Vol. 63: 1379-1389.

Sen, Z. 1979. Application of the Autorun Test to Hydrologic Data. Journal of Hydrology. Vol. 42: 1-7.

Sen, Z. 1982. Discussion of Statistical Considerations and Sampling Techniques for Ground-Water Quality Monitoring by J.D. Nelson and R.C. Ward. Vol. 20: 494-495.

Sen, Z. 1987. Discussion of Statistical Methods for Characterizing Ground-

Water Quality by Harris et al. Ground Water. Vol. 25, No. 6 (November-December): 741-742.

Shapiro, S.S. and M.B. Wilk. 1965. An Analysis of Variance Test for Normality (complete samples). Biometrika. Vol. 52: 591-611.

Shapiro, S.S., M.B. Wilk and Mrs. H.J. Chen. 1968. A Comparative Study of Various Tests for Normality. American Statistical Assoc. Journal (December): 1343-1372.

Silver, Carl A. 1986a. Statistical Approaches to Groundwater Monitoring. Open File Report #7. University of Alabama, Environmental Institute for Waste Management Studies.

Silver, Carl A. 1986b. Statistical Approaches to Ground Water Monitoring. Journal of Hazardous Materials. Vol. 13: 207-216.

Silver, Carl A. and Dennis Dunn. 1988. Statistical Analysis of Rare Events in Ground Water. Journal of Hazardous Materials. Vol. 18: 17-24.

Snedecor, G.W. and W.G. Cochran. 1967. Statistical Methods. Ames, Iowa: Iowa State University Press.

Songzoni, William. September, 1988. Personal Communication. Wisconsin State Lab of Hygiene.

Sugden, D.E. and B.S. John. 1984. Glaciers and Landscape. London: Butler & Tanner, Ltd.

Tusler, Mark. September, 1988. Personal Communication. Wisconsin Department of Natural Resources, Bureau of Solid and Hazardous Waste Management.

vanBelle, Gerald and James P. Hughes. 1984. Nonparametric Tests for Trend in Water Quality. Water Resources Research. Vol. 20 (January): 127-136.

Wisconsin DNR. 1985. Chapter NR 140 Wisconsin Administrative Code. Cr. Register, September, 1985, No. 357 and as amended Cr. Register, October, 1988, No. 394.

Zaporozex, A. and R.D. Cotter. 1885. Major Ground-Water Units of Wisconsin. University of Wisconsin Extension: Geological and Natural History Survey. Madison, WI.

Zar, Jerrold H. 1982. Power and Statistical Significance in Impact Evaluation. Ground Water Monitoring Review. Vol. 2, No. 3: 33-35.

APPENDIX A

Recommended Statistical Tests

APPENDIX A
RECOMMENDED STATISTICAL TESTS

TABLE OF CONTENTS

LIST OF TABLES

APPENDIX A
RECOMMENDED STATISTICAL TESTS

Three statistical procedures are documented below. The tests are 1) the skewness test for normality, 2) the Mann-Kendall test for trend, and 3) calculation of prediction intervals. The documentation provided here is sufficient to perform the tests; however, the reader is cautioned against applying statistical procedures without thorough understanding of the theory, hypotheses and limitations of the procedures. The references cited can provide the necessary background information.

Many computer based statistical analysis programs include the skewness test (or calculation of the skewness coefficient) and the Mann-Kendall test for trend (or calculation of Kendall's Tau statistic).

A1.0 The Skewness Test

References:

George W. Snedecor and William G. Cochran. *Statistical Methods.*The Iowa State University Press. Ames, Iowa. 1980.

Harris et al. Statistical Methods for Characterizing Ground-Water Quality. *Groundwater*, Vol.25, No.2. March-April 1987.

The skewness test may be used to determine whether or not a set of independent data points are drawn from a normal distribution. The test is very simple to apply. The skewness coefficient is calculated and the value compared to a critical value found in Table A-1.

The null and alternative hypotheses for the skewness test may be stated as:

H_0 : The data may be normally distributed.

H_1 : The data are not drawn from a normal distribution.

The skewness coefficient is:

$$g = \frac{\sqrt{n} \sum_{i=1}^{n} (x_i - \overline{x})^3}{\sum_{i=1}^{n} (x_i - \overline{x})^{\frac{3}{2}}}$$

where g = skewness coefficient
n = sample size
x_i = concentration at time i
$\overline{X}$ = mean concentration

To apply the skewness test, consult Table A-1. Find the critical skew associated with sample size n. If the calculated skewness coefficient is **less** than the critical skew (from the table), then the null hypothesis **is not rejected** at the 5 or 1 percent significance level. Conclude that the data may be drawn from a normal distribution. If the calculated skewness coefficient is **greater** than the critical skew in the table, then the null hypothesis **is rejected** at the 5 or 1 percent significance level. Conclude that the data are probably not drawn from a normal distribution.

Note that Table A-1 may be used to test either positive skew, negative skew or both. To test negative skew, the critical skew is just the negative of the tabulated values at any sample size. To perform a two-tailed test, compare the calculated skewness coefficient to both the positive and negative critical skew value. The test is then performed at the 10 or 2 percent significance level.

TABLE A-1
CRITICAL SKEWNESS COEFFICIENTS

Sample Size	Percentage Points 5%	Percentage Points 1%
9	0.953	1.420
10	0.950	1.395
11	0.927	1.358
12	0.915	1.331
13	0.886	1.306
14	0.861	1.291
15	0.854	1.280
16	0.833	1.246
17	0.817	1.220
18	0.798	1.197
19	0.769	1.161
20	0.777	1.146
21	0.753	1.116
22	0.742	1.099
23	0.732	1.087
24	0.710	1.074
25	0.712 (0.711)	1.060 (1.061)
26	0.689	1.013
27	0.689	1.016
28	0.674	1.006
29	0.669	0.992
30	0.651 (0.662)	0.972 (0.986)
35	(0.621)	(0.921)
40	(0.587)	(0.869)
45	(0.558)	(0.825)
50	(0.533)	(0.787)

Adapted from Table 4 (Harris et al, 1987) . Data in parentheses from Table A-20 (Snedecor and Cochran 1980)

A2.0 Mann-Kendall Test

The Mann-Kendall test for trend is also known as Kendall's Tau test. The one sided null hypothesis is:

H_O = the X_i exhibit no trend

The one sided alternative hypotheses are:

H_+ = the X_i exhibit an upward trend

H_- = the X_i exhibit a downward trend

and the two-sided alternative is:

H = there is either an upward or a downward trend.

A good reference for Kendall's Tau statistic and the associated statistical hypothesis test is

Gibbons, Jean Dickenson. *Nonparametric Methods for Quantitative Analysis (Second Edition)* . American Sciences Press, Inc.: Columbus, Ohio. 1985.

The test for trend assumes that the data points are independent of each other. This would not be the case for monthly data, for example, if there is a seasonal pattern by months. If the seasonal variation can be removed, these procedures are applicable to the adjusted data. Kendall's Tau statistic is the nonparametric analog of the parametric test based on the regression coefficient (which assumes the normal distribution) . The asymptotic efficiency of Kendall's tau to the regression test is about 0.98 for normal distributions.

To perform the test, consider each sampling event as a pair of observations (X,Y) where Y is the sample value and X is the sampling date. List the data in chronological order and assign ranks to X and to Y independently. Rank (1) is associated with the first sampling date, X_1, and the lowest sample value (if testing for positive trend).

If the X and the Y characteristics are in perfect agreement (positive trend), the Y data should be in natural order (the X data already are). If there is perfect disagreement (negative trend), the corresponding Y data is in reverse of natural order. The Kendall Tau coefficient is a relative measure of the discrepancy between the actual observed order of the Y's and the two orders that would result from perfect association. The procedure is most easily explained by an example (drawn from Gibbons 1985):

EXAMPLE:

Suppose that n=5 and two sets of ranks are paired as follows.

X rank: 1 2 3 4 5 (monitoring dates)

Y rank: 2 3 1 4 5 (sample values)

Note that the X set is in natural order. In the resulting arrangement of Y ranks, we consider all of the possible pairs of Y ranks and score a 1 for each pair of ranks that appear in natural order and -1 for those in reverse order. We take the pairs in a systematic way, as the 2 paired with each successive rank appearing to its right, then the 3 paired with each to its right, and so on. The first pair of Y ranks, 2 followed by 3, is in natural order, so its score is 1. The second pair, 2 followed by 1, is in reverse order, so -1 is scored. The resulting scores for all possible pairs are shown in Table A-2.

Note that there are $\binom{5}{2} = 10$ possible pairs. The ratio of the total plus score, in this case 8, to the maximum,10, provides a measure of relative agreement, that is 8/10. Similarly the ratio of the total minus score to the maximum, 2/10 in this case, measures the relative disagreement. The net relative score of association is then 8/10-2/10 = 6/10, and this is the value of the Kendall Tau statistic. If we let U= the number of pairs of Y values (or ranks) in natural order (that is, the number of plus scores) and let V equal the number of Y pairs in reverse order, and let S be the difference between U and V, S = U-V , then the Tau coefficient is calculated as:

$$T = \frac{2S}{n\,(n-1)}$$

where n is the number of (X,Y) pairs (sample size). For our example, T is equal to 3/5 (0.60). The test statistic presented here is not exact if ties exist in the sample data. No ties should exist in the X series, sampling dates. Any duplicate data should be averaged. If multiple dates have the same sample value the denominator of the test statistic must be adjusted (i.e. ties exist). The adjusted test statistic is more difficult to calculate and computer analysis is recommended. Gibbons (1985) presents this test statistic in detail and how it may be calculated by hand.

TABLE A-2
Calculation of Kendall Tau Statistic

Y pair	Score	Summary Totals
2,3	1	8 plus
2,1	-1	2 minus
2,4	1	
2,5	1	
3,1	-1	
3,4	1	
3,5	1	
1,4	1	
1,5	1	
4,5	1	

The statistic T above, and the T statistic adjusted for ties, is found in most statistical software packages. Interpretation of the Tau statistic is simple. If perfect positive correlation exists T is equal to 1. If perfect negative correlation exists T is equal to -1. If no correlation exists T is equal to 0. To test the null hypothesis that no trend exists, the value of T is compared to a critical value of T found on Table A-3.

For a one-sided test for positive trend, compare the calculated value of T to the associated T in Table A-3. If the calculated value is **greater than or equal to** the table value, reject the null hypothesis of no trend (conclude that positive trend exists). For a one-sided test for negative trend, compare the calculated value of T to the negative of the associated T in Table A-3. If the calculated value is **less than or equal to** the table value, reject the null hypothesis of no trend (conclude that negative trend exists). For a two-sided test for positive or negative trend, compare the calculated value of T to both the positive and negative of the associated table value. If the calculated value is greater than +T or less than -T conclude that trend exists.

TABLE A-3
Critical values for Kendall Tau Statistic

Sample Size (n)	T (tau)
5	0.80
6	0.733
7	0.619
8	0.571
9	0.500
10	0.467
11	0.418
12	0.394
13	0.359
14	0.363
15	0.333
16	0.317
17	0.309
18	0.294
19	0.287
20	0.274
21	0.267
22	0.264
23	0.257
24	0.246
25	0.24
26	0.237
27	0.231
28	0.228
29	0.222
30	0.218

Note: The T values for $n>10$ are the right tail (or left tail) critical values for a one-sided test performed at a significance level of 0.05. For $5<n<10$, the values are the lowest T for which a one sided test, performed at a significance level of 0.05, would reject the null hypothesis. In this case, the probability associated with the T value is not exact, but is always less than 0.05. For n>30, critical values may be found from a normal probability table (See Gibbons 1985). However, T = 0.218 will always be a conservative estimate.

A3.0 PREDICTION INTERVALS

References:

U.S. Environmental Protection Agency. *Statistical Analysis of Groundwater at RCRA Facilities*. Office of Solid Waste, Waste Management Division. October, 1988 (Available from NTIS Reference Number PB 89-151-047).

Gibbons, Robert D. "Statistical Prediction Intervals for the Evaluation of Ground-Water Quality." *Ground Water*. Vol. 25. pp.455-465. 1987.

A prediction interval is a statistical interval designed to define a background concentration interval within which future measurements from the same population are likely to fall. The prediction interval can answer the question "What is the concentration associated with an allowable exceedance probability given the natural variability in the data and the sample size?". The allowable exceedance probability is recommended as 0.05 by EPA.

The prediction interval is recommended to be developed on a well by well basis. Data from multiple wells should not be aggregated.

To calculate a prediction interval the mean, $\bar{x}$, and the standard deviation s, must be calculated for the data used to form the prediction interval. Then the interval is given by

$$\bar{x} + s\sqrt{\frac{1}{m}+\frac{1}{n}}\ t_{(n-1,K,0.95)}$$

where m is the number of measurements per sampling period (i.e. 2 if duplicate data are available), and n is the number of observations in the background data, and $t_{(n-1,K,\,0.95)}$ is found from Table A-4. The table is entered with K as the number of future observations (usually 1 if comparison is done each quarter, or 4 if comparison is done annually), and degrees of freedom, ν = n-1. If K is greater than 5 (unlikely), use the column for K = 5.

To compare new data to the prediction interval, calculate the mean of duplicate measurements or just compare the new data point to see whether it falls within the interval. If the new data is not within the prediction interval, this is statistically significant evidence of contamination.

Note that for a single future observation (i.e. one observation per quarter with quarterly comparisons), the t value may be obtained straight from the t-distribution which is tabulated in most statistical texts. Also, note that the prediction intervals are one-sided, giving a value that should not be exceeded by the future observations. If a two sided interval is required, the same procedure may be used, however Table A-4 will provide interval at the 2 α percent significance level (where α is usually 0.05).

TABLE A-4
95th Percentiles of the Bonferroni t-statistics, $t_{(\nu, \alpha/k)}$
(adapted from EPA, October 1988)

	k	1	2	3	4	5
ν	α/k	0.05	0.025	0.0167	0.0125	0.01
4		2.13	2.78	3.20	3.51	3.75
5		2.02	2.57	2.90	3.17	3.37
6		1.94	2.45	2.74	2.97	3.14
7		1.90	2.37	2.63	2.83	3.00
8		1.86	2.31	2.55	2.74	2.90
9		1.83	2.26	2.50	2.67	2.82
10		1.01	2.23	2.45	2.61	2.76
15		1.75	2.13	2.32	2.47	2.60
20		1.73	2.09	2.27	2.40	2.53
30		1.70	2.04	2.21	2.34	2.46
>30		1.65	1.96	2.13	2.24	2.33

ν = degrees of freedom associated with the mean square error.
k = number of comparisons
α = 0.05, the experimentwise error level

APPENDIX B

Control Well Groups

APPENDIX B

CONTROL GROUP SITES

SITE	LICENSE
City of Janesville	2822
City of Medford	341
City of Merrill	912
City of Oconto	137
City/Town of Cedarburg	271
County Dane #1 - Verona	2680
County Eau Claire Seven Mile Creek	2821
County Lacrosse	2637
County Marathon	2892
County Portage	2966
County Sauk (Old)	2051
County Sauk (New)	2978
Fort Howard Paper Co. Green Bay	2332
Rock County / City of Janesville	3023
Town of Washington	160
Village of Bonduel	59
Waste Management Inc Greidanus Landfill	140
Waste Management Inc. Brookfield Landfill	1
Wausau Paper Mills	2875
Wisconsin Electric Power - Oak Creek	2357

CONTROL GROUP WELLS BY LICENSE NUMBER

LICENSE	DNR WELL ID	WELL GROUP
1	1	3
1	2	3
1	3	3
1	4	4
1	5	4
1	6	4
1	7	4
1	8	1
1	11	4
1	12	4
1	13	1
1	14	3
1	15	3
1	16	4
1	19	1
59	1	1
59	2	4
59	3	3
59	4	3
59	5	3
59	6	4
137	1	4
137	2	4
137	3	1
137	4	3
137	5	1
137	9	1
137	10	1
137	11	3
137	12	1
137	13	1
137	14	2
137	15	3
137	16	4
137	17	2
140	202	4
140	203	1
140	204	1
140	205	2
140	206	2
140	207	3
140	208	3
140	211	1
140	212	1
140	213	1
140	214	3
140	224	1
140	225	1

CONTROL GROUP WELLS BY LICENSE NUMBER

LICENSE	DNR WELL ID	WELL GROUP
160	101	1
160	102	4
160	103	1
160	104	1
160	105	3
160	107	4
271	201	1
271	202	3
271	209	2
271	210	1
271	211	4
271	212	2
271	213	2
271	216	3
271	217	4
271	218	4
271	219	2
341	801	1
341	802	4
341	803	2
341	804	1
341	805	1
341	806	2
341	807	1
341	808	4
341	809	4
341	810	4
341	811	4
341	812	1
912	2	4
912	6	1
912	7	3
912	8	3
912	9	4
912	15	4
912	16	4
912	17	1
912	18	1
912	19	1
912	20	4
912	21	1
912	22	4
912	23	3
912	24	1
912	25	3
912	26	2
912	27	2
2051	104	1

CONTROL GROUP WELLS BY LICENSE NUMBER

LICENSE	DNR WELL ID	WELL GROUP
2051	105	1
2051	106	4
2051	110	1
2051	114	4
2051	115	4
2051	116	4
2051	117	4
2051	118	2
2051	119	3
2051	120	2
2051	121	4
2051	122	4
2051	123	3
2051	107	1
2332	1	2
2332	4	2
2332	5	4
2332	6	4
2332	7	1
2332	8	4
2332	9	1
2332	10	2
2332	11	4
2332	12	2
2332	13	1
2332	14	3
2332	15	1
2332	16	4
2332	17	1
2332	18	1
2332	19	1
2332	20	4
2332	21	1
2332	22	1
2332	23	4
2332	32	4
2332	33	1
2357	201	3
2357	202	1
2357	203	1
2357	204	4
2357	205	1
2357	206	1
2357	207	2
2357	208	3
2357	209	2
2357	210	2
2357	211	3

CONTROL GROUP WELLS BY LICENSE NUMBER

LICENSE	DNR WELL ID	WELL GROUP
2357	212	4
2357	213	4
2357	214	4
2357	227	4
2357	228	4
2357	229	3
2637	4	1
2637	6	2
2637	7	2
2637	8	2
2637	9	2
2637	10	1
2637	11	1
2637	12	1
2637	13	3
2637	14	1
2637	15	1
2637	16	4
2680	106	1
2680	108	2
2680	114	4
2680	115	2
2680	124	3
2680	125	1
2680	126	4
2680	131	4
2680	134	1
2680	135	2
2680	136	1
2680	140	1
2680	150	1
2680	171	3
2680	172	4
2680	173	3
2680	175	4
2680	176	3
2680	177	1
2680	178	4
2680	179	4
2680	180	4
2680	181	1
2821	5	1
2821	6	1
2821	7	3
2821	8	1
2821	17	3
2821	18	4
2821	19	4

CONTROL GROUP WELLS BY LICENSE NUMBER

LICENSE	DNR WELL ID	WELL GROUP
2821	30	1
2822	107	4
2822	108	4
2822	109	1
2822	110	1
2822	112	4
2822	113	4
2822	114	2
2822	115	2
2822	124	4
2822	125	4
2822	129	4
2875	1	4
2875	2	1
2875	3	3
2875	4	1
2875	5	4
2875	9	1
2875	10	1
2875	11	1
2892	1	1
2892	2	1
2892	8	1
2892	10	1
2892	34	1
2892	35	1
2892	36	3
2892	37	2
2892	38	1
2966	1	1
2966	2	1
2966	4	2
2966	5	2
2966	9	1
2966	10	1
2966	12	1
2966	13	1
2966	14	1
2966	16	1
2966	17	1
2966	23	2
2966	24	2
2966	26	1
2966	27	1
2966	28	1
2966	30	1
2966	31	1
2966	32	4

CONTROL GROUP WELLS BY LICENSE NUMBER

LICENSE	DNR WELL ID	WELL GROUP
2966	33	1
2966	29	1
2978	101	1
2978	102	1
2978	103	1
2978	104	1
2978	106	1
2978	107	1
2978	108	1
2978	111	1
2978	118	1
2978	119	1
2978	105	1
2978	109	1
3023	1	2
3023	2	1
3023	3	1
3023	4	4
3023	5	2
3023	6	1
3023	7	1
3023	8	1
3023	9	4
3023	10	4
3023	11	2
3023	12	3
3023	13	4
3023	14	4
3023	15	4
3023	16	1

APPENDIX C

Statistical Screening Results

APPENDIX C

The results of the statistical screening of the Wisconsin grounwater quality database are presented herein. The screening results are presented first alphabetically by site name and then by the predictor groups.

The column headings refer to:

# WELLS PREDICTED	:	This is the number of wells to which the predictor was applied. That is, it is the number of wells *with enough data* at the site. "Enough data" is defined as 8 or more sampling dates for at least two parameters at the well.
TOTAL # WELLS	:	This is the total number of wells at the site, with or without enough data.
% TOTAL DIRTY	:	This is the number of wells with enough data which are predicted to be "dirty" divided by the total number of wells.

Alphabetical Site List

FACILITY NAME	LICENSE	# WELLS TESTED	TOTAL # WELLS	% TOTAL DIRTY
ANDERSON PEAT-ORGANIC COMPOST	420	7	7	57.14
APPLETON PAPERS, INC	30	18	18	88.89
BAAP DETERRENT BURNING GROUND	3037	4	4	50
BAAP-PROPELLANT BURNING GRNDS	2814	7	7	28.57
BADGER ARMY AMMUNITION PLANT	2813	6	6	50
BADGER DISPOSAL	234	12	12	33.33
BAKER SANITARY LANDFILL	189	4	4	25
BARRETT LANDFILL, INC	1940	27	27	92.59
BEECHER WOODYARD	2328	3	3	33.33
BELOIT CONCRETE STONE CO INC	781	2	2	100
BERGSTROM PAPER LF-NEENAH	2446	30	30	96.67
BRIDGEPORT LANDFILL	445	3	3	0
BRILLION IRON WORKS, INC	2866	6	6	66.67
CENTRAL SANITARY LANDFILL	2132	8	8	75
CNTY BROWN-EAST	2569	35	35	48.57
CNTY BROWN-WEST	2568	35	35	71.43
CNTY DANE LANDFILL #1-VERONA	2680	22	22	72.73
CNTY DANE LANDFILL #2-RODEFELD	3018	32	32	59.38
CNTY DOOR SANITARY LANDFILL	2937	10	10	80
CNTY EAU CLAIRE-SEVENMILE CRK	2821	11	11	0
CNTY FOND DU LAC	2358	18	18	88.89
CNTY GREEN	217	9	9	55.56
CNTY GREEN S/W DISPOSAL SITE	2990	10	10	80
CNTY JUNEAU	2565	3	3	0
CNTY KEWAUNEE SW BALEFILL	2975	11	11	100
CNTY LA CROSSE	2637	11	11	9.09
CNTY MARATHON LANDFILL	2892	7	7	14.29
CNTY MILWAUKEE HWY DEPT	881	11	11	90.91
CNTY MONROE-RIDGEVILLE SITE	2858	6	6	0
CNTY ONEIDA	2805	4	4	0
CNTY OUTAGAMIE	2484	33	33	42.42
CNTY PORTAGE LANDFILL	2966	20	20	5
CNTY SAUK	2051	15	15	20
CNTY SAUK SANITARY LANDFILL	2978	11	11	0
CNTY SHAWANO-ANGELICA SITE	2728	2	2	50
CNTY WINNEBAGO	611	78	78	76.92
COLT INDUSTRIES-FARNAM DIV.	640	4	4	25
CONSOLIDATED PAPER WIS RIV DIV	1686	22	27	45.45
CONSOLIDATED PAPERS-BIRON DIV	1687	8	8	75
CONSOLIDATED PAPERS-KRAFT DIV	1838	16	16	100
CONSOLIDATED PAPERS-STEVENS PT	2344	24	24	29.17
CONSOLIDATED PAPERS-WQC	2488	33	33	87.88
CTY ABBOTSFORD LANDFILL	2932	5	5	20
CTY ADAMS-VIL FRIENDSHIP	1721	3	3	33.33
CTY ALGOMA	179	6	6	33.33
CTY ANTIGO	1357	15	15	26.67
CTY ASHLAND	177	21	21	95.24

Alphabetical Site List

FACILITY NAME	LICENSE	# WELLS TESTED	TOTAL # WELLS	% TOTAL DIRTY
CTY BARRON	82	2	2	50
CTY BLACK RIVER FALLS	287	10	10	50
CTY BURLINGTON	186	5	5	100
CTY CHIPPEWA FALLS	85	7	7	71.43
CTY CLINTONVILLE	314	10	10	50
CTY EAU CLAIRE	77	3	3	33.33
CTY FOX LAKE WOODBURNING SITE	369	4	4	100
CTY GALESVILLE	2738	1	1	0
CTY GILLETT	1115	5	5	40
CTY GREEN BAY-2130 DANZ AVENUE	170	6	6	66.67
CTY GREEN BAY-HUMBOLT ROAD	1129	6	6	50
CTY GREEN BAY-MILITARY AVENUE	169	5	5	100
CTY HAYWARD	1751	2	2	50
CTY JANESVILLE	62	7	7	57.14
CTY JANESVILLE	2822	10	10	70
CTY JANESVILLE-ASH BEDS	3061	5	5	60
CTY KENOSHA	38	5	5	100
CTY LA CROSSE	144	5	5	100
CTY MADISON-GREENTREE HILLS	1714	5	5	100
CTY MADISON-SYCAMORE SITE	1935	7	7	85.71
CTY MEDFORD	341	11	11	63.64
CTY MENOMONIE	372	5	5	60
CTY MERRILL	912	17	17	29.41
CTY NEW RICHMOND	2492	3	3	0
CTY OCONTO	137	14	14	57.14
CTY PARK FALLS	777	1	1	100
CTY PHILLIPS	57	2	2	0
CTY PORTAGE	1885	6	7	33.33
CTY RICE LAKE	108	2	2	100
CTY RICHLAND CENTER	1519	4	4	50
CTY RIPON-TN RIPON	467	3	3	100
CTY SHAWANO	2342	17	17	76.47
CTY SHAWANO - PHASE 2	3069	12	12	8.33
CTY STOUGHTON	133	5	5	80
CTY SUPERIOR-MOCASSIN MIKE	2627	4	4	100
CTY SUPERIOR-WIS POINT LF	12	12	12	100
CTY TWO RIVERS	318	5	5	100
CTY WATERTOWN	893	3	3	100
CTY WAUPUN	2246	4	4	100
CTY WEST BEND	224	18	18	61.11
CTY WHITEWATER	65	9	9	77.78
CTY WISCONSIN DELLS	2712	8	8	0
CTY-TN CEDARBURG	271	11	11	36.36
DAIRYLAND POWER COOP-ALMA	1673	12	12	33.33
DAIRYLAND POWER COOP-CASSVILLE	96	25	25	28
DAIRYLAND POWER COOP-GENOA #3	1747	6	6	100
DAIRYLAND POWER-OFFSITE DISP.	2927	13	13	61.54

Alphabetical Site List

FACILITY NAME	LICENSE	# WELLS TESTED	TOTAL # WELLS	% TOTAL DIRTY
DEROSSO LANDFILL	1979	5	5	40
EXXON MINERALS COMPANY	2977	13	13	0
FAHERTY DRILLING CO INC	949	2	2	50
FALK CORPORATION	1882	8	8	62.5
FLAMBEAU PAPER CORP	2756	13	13	76.92
FORT HOWARD PAPER CO-GREEN BAY	2332	24	24	62.5
FORT HOWARD STEEL & WIRE DIV	2972	7	7	100
GENERAL MOTORS-WHEELER PIT	2795	8	8	75
GREDE-REEDSBURG FOUNDRY SW LF	2974	6	6	16.67
GREEN LAKE SANITARY LANDFILL	1890	5	5	60
H & R PAPER & REFUSE SERVICE	850	12	12	33.33
HOLTZ & KRAUSE, INC	674	12	12	75
HUGHES REFUSE & LANDFILL CO	2776	5	5	0
JACKSON CNTY SANITARY LF INC C	2004	5	5	0
JACKSON COUNTY IRON COMPANY	2924	7	7	57.14
JAMES RIVER CORP-ASHLAND MILL	2826	3	3	33.33
JAMES RIVER NORWALK-ALPINE	1832	4	4	100
JAMES RIVER NORWALK-NORTHLAND	2893	11	11	100
JONGETJES LANDFILL	943	4	4	75
JUNKER SANITARY LANDFILL, INC	1972	2	2	50
KIMBERLY-CLARK LAKEVIEW MILL	3004	7	7	57.14
KOHLER COMPANY LANDFILL	1508	16	16	18.75
LAKE AREA DISPOSAL LANDFILL	2054	5	5	40
LAND AND GAS RECLAMATION, INC	1118	4	4	75
LAND RECLAMATION, LTD	572	28	28	60.71
LAWENT IRON & METAL CORP	2611	5	5	20
LEADFREE LANDFILL-BRIDGEPORT	2959	2	2	0
MADISON PRAIRIE DEMOLITION LF	2918	11	11	27.27
MASTER DISPOSAL, INC LANDFILL	2425	7	7	100
MAZO LAND DISPOSAL	2009	2	2	50
MERCURY MARINE LANDFILL	2603	1	1	0
METROPOLITAN REFUSE DIST, INC	107	14	14	57.14
MIDWEST DISPOSAL	73	15	15	66.67
MOSINEE PAPER CORP. LANDFILL	2806	22	22	45.45
MURRAY MACHINERY, INC	1722	1	1	100
N.O.W. PAPER CORP. FLY ASH LF	2964	2	2	0
NEENAH-WHITING MILL LANDFILL	2576	6	6	50
NEKOOSA MILL REFUSE DISP SITE	2857	8	8	100
NEKOOSA PAPER-WW TREATMENT RES	2613	38	38	65.79
NEKOOSA PAPERS (LIME SLUDGE)	2614	14	14	85.71
NEKOOSA PAPERS ASH-BARK SITE	1365	7	7	71.43
NEKOOSA PAPERS INC	2891	3	3	0
NEKOOSA PAPERS, INC	2811	6	6	0
NEKOOSA PAPERS,SLUDGE SPREAD.	2672	6	6	66.67
NIAGARA OF WISC PAPER CORP	3005	2	2	50
NORTH WOODS DISPOSAL	2001	6	6	66.67
NORTHRNESTATESKPWR-DEERPCREEK	2767	2	2	0

Alphabetical Site List

FACILITY NAME	LICENSE	# WELLS TESTED	TOTAL # WELLS	% TOTAL DIRTY
OCEAN SPRAY CRANBERRIES, INC	2423	10	10	30
OWENS-ILLINOIS, INC LANDFILL	1346	32	40	56.25
PATS STORAGE LAGOON	3003	3	3	0
PELISHEK CONTRACTING LANDFILL	338	1	1	100
POPE & TALBOT WI-ABSORBENT PRD	2695	13	13	23.08
R L O'KEEFE & SONS, INC LF	2031	6	6	50
REFUSE HIDEAWAY LANDFILL	1953	4	4	100
RHINELANDER PAPER COMPANY	1857	22	23	45.45
RICHLAND CENTER FOUNDRY CO	2487	6	6	66.67
RIVERSIDE SANITATION LANDFILL	738	2	2	0
ROCK COUNTY-CTY JANESVILLE LF	3023	15	15	33.33
RUEF SANITARY LANDFILL	2936	5	5	100
RUEF'S SANITARY SERVICE, INC	478	5	5	60
SANITARY TRANS & LF-DELAFIELD	719	51	52	76.47
SCOTT PAPER CO	2368	6	6	50
SCOTT PAPER CO LANDFILL	2846	4	4	0
SHAWANO PAPER MILLS LANDFILL	2719	6	6	50
SLINGER FOUNDRY LANDFILL	2702	3	3	100
SOUTHEASTERN BARRON CNTY	1887	2	2	100
THILMANY PULP & PAPER CO	493	16	16	31.25
TN ASHWAUBENON	263	4	4	75
TN EAST TROY	24	20	20	40
TN GRAND RAPIDS	693	2	2	0
TN HALLIE	2807	2	2	0
TN LINCOLN	1779	2	2	50
TN MENASHA	671	4	4	50
TN MENOMONIE	2659	2	2	0
TN MINOCQUA BO-DI-LAC LANDFILL	1561	2	2	50
TN MINOCQUA-HWY 51 SITE	1558	3	3	33.33
TN MINOCQUA-MERCER LAKE SITE	1559	3	3	0
TN ONALASKA	507	9	9	55.56
TN ROSENDALE	2747	2	2	50
TN RUTLAND	2115	2	2	50
TN SHERMAN	2856	2	2	0
TN ST GERMAIN	1389	3	3	33.33
TN STUBBS-DISTRICT 5 LANDFILL	2909	1	1	0
TN SUGAR CAMP-SOUTH SITE	2884	1	1	0
TN WASHINGTON	160	5	5	40
TN WESCOTT	1004	4	4	50
TOMAHAWK TISSUE CORP LANDFILL	1878	10	10	80
TORK ALUM MUD DISPOSAL SITE	1892	12	12	91.67
TORK LANDFILL CORP (SENECA)	2967	17	17	29.41
TORK LANDFILL CORPORATION	652	16	29	93.75
U S ARMY-BAAP ACID SPILL AREA	2934	6	6	83.33
U S ARMY-BAAP PERIMETER WELLS	3038	7	7	100
VALLEY SANITATION CO, INC	2686	3	3	0
VAN HANDEL SANITARY LANDFILL	49	10	10	90

Alphabetical Site List

FACILITY NAME	LICENSE	# WELLS TESTED	TOTAL # WELLS	% TOTAL DIRTY
VIL ROTHSCHILD	1538	2	2	50
VIL WONEWOC	164	4	4	75
VULCAN MATERIALS CO	2998	15	23	93.33
WARD PAPER COMPANY LANDFILL	2991	7	8	0
WASTE CONTROL INC	1970	6	6	16.67
WASTE MANAGEMENT OF GREEN BAY	3	11	11	100
WASTE MANAGEMENT OF WIS-NEOSHO	443	8	8	62.5
WASTE MGMT OF WIS-GREIDANUS LF	140	10	10	20
WASTE MGMT OF WIS-METRO LF	1099	42	42	64.29
WASTE MGMT OF WIS-MUSKEGO LF O	2895	4	4	50
WASTE MGMT OF WIS-OMEGA HILLS	1678	109	109	63.3
WASTE MGMT OF WIS-PHEASANT RUN	1739	16	17	43.75
WASTE MGMT OF WIS-RIDGE VIEW	2575	16	16	43.75
WASTE MNGT OF WI, INC-POLK	307	14	14	92.86
WASTE MNGT OF WIS, INC-CITY DS	37	12	12	100
WASTE MNGT OF WIS, INC-MUSKEGO	141	10	10	90
WASTE MNGT OF WIS, INC-RECLAM	1356	5	5	100
WASTE MNGT OF WIS-BROOKFIELD	1	14	14	85.71
WATERFORD SEPTIC SERVICE	2894	2	2	0
WAUPACA FOUNDRY COMPANY	2638	4	4	75
WAUPACA FOUNDRY, INC	2089	1	1	0
WAUSAU HOMES, INC	1774	4	4	25
WAUSAU PAPER MILLS LANDFILL	2875	7	7	28.57
WEYERHAEUSER COMPANY	2873	9	9	11.11
WI DOT HWY 100-RYAN RD	2988	7	7	71.43
WI ST DEPT TRANSPORTATION	2586	2	2	50
WIS ELEC POWR-PLEASANT PRAIRIE	2786	11	11	36.36
WIS ELECTRIC POWER CO-HWY 32	2801	11	11	9.09
WIS ELECTRIC POWER CO-HWY 59	918	1	1	0
WIS ELECTRIC POWER-CEDAR SAUK	603	30	37	33.33
WIS ELECTRIC POWER-OAK CREEK	2357	17	17	52.94
WIS POWER & LIGHT CO-COLUMBIA	2325	6	15	33.33
WIS POWER & LIGHT CO-EDGEWATER	2524	6	8	50
WIS POWER & LIGHT-NELSON DEWEY	2525	17	17	17.65
WIS POWER & LIGHT-ROCK RIVER	728	5	5	40
WIS PUB SERV CORP-WESTON #3 LF	2879	8	8	0
WIS PUBLIC SERV CORP-LEGNER	3067	1	1	0
WIS PUBLIC SERV-FLY ASH SITE	51	5	5	100
WP&L EDGEWATER GEN STA-DRY ASH	2853	11	11	72.73
YOURCHUCK'S SANITARY LANDFILL	2010	2	2	0

SITE LIST BY PREDICTOR GROUPS

FACILITY NAME	LICENSE	# WELLS PREDICTED	TOTAL # WELLS	% TOTAL DIRTY
WIS ELECTRIC POWER CO-HWY 59	918	1	1	0
WAUPACA FOUNDRY, INC	2089	1	1	0
MERCURY MARINE LANDFILL	2603	1	1	0
CTY GALESVILLE	2738	1	1	0
TN SUGAR CAMP-SOUTH SITE	2884	1	1	0
TN STUBBS-DISTRICT 5 LANDFILL	2909	1	1	0
WIS PUBLIC SERV CORP-LEGNER	3067	1	1	0
CTY PHILLIPS	57	2	2	0
TN GRAND RAPIDS	693	2	2	0
RIVERSIDE SANITATION LANDFILL	738	2	2	0
YOURCHUCK'S SANITARY LANDFILL	2010	2	2	0
TN MENOMONIE	2659	2	2	0
NORTHRNESTATESKPWR-DEERPCRE	2767	2	2	0
TN HALLIE	2807	2	2	0
TN SHERMAN	2856	2	2	0
WATERFORD SEPTIC SERVICE	2894	2	2	0
LEADFREE LANDFILL-BRIDGEPORT	2959	2	2	0
N.O.W. PAPER CORP. FLY ASH LF	2964	2	2	0
BRIDGEPORT LANDFILL	445	3	3	0
TN MINOCQUA-MERCER LAKE SITE	1559	3	3	0
CTY NEW RICHMOND	2492	3	3	0
CNTY JUNEAU	2565	3	3	0
VALLEY SANITATION CO, INC	2686	3	3	0
NEKOOSA PAPERS INC	2891	3	3	0
PATS STORAGE LAGOON	3003	3	3	0
CNTY ONEIDA	2805	4	4	0
SCOTT PAPER CO LANDFILL	2846	4	4	0
JACKSON CNTY SANITARY LF INC C	2004	5	5	0
HUGHES REFUSE & LANDFILL CO	2776	5	5	0
LAWENT IRON & METAL CORP	2611	5	5	20
CTY ABBOTSFORD LANDFILL	2932	5	5	20
BAKER SANITARY LANDFILL	189	4	4	25
COLT INDUSTRIES-FARNAM DIV.	640	4	4	25
WAUSAU HOMES, INC	1774	4	4	25
CTY EAU CLAIRE	77	3	3	33.33
TN ST GERMAIN	1389	3	3	33.33
TN MINOCQUA-HWY 51 SITE	1558	3	3	33.33
CTY ADAMS-VIL FRIENDSHIP	1721	3	3	33.33
BEECHER WOODYARD	2328	3	3	33.33
JAMES RIVER CORP-ASHLAND MILL	2826	3	3	33.33
TN WASHINGTON	160	5	5	40
WIS POWER & LIGHT-ROCK RIVER	728	5	5	40
CTY GILLETT	1115	5	5	40
DEROSSO LANDFILL	1979	5	5	40
LAKE AREA DISPOSAL LANDFILL	2054	5	5	40
CTY BARRON	82	2	2	50
FAHERTY DRILLING CO INC	949	2	2	50
VIL ROTHSCHILD	1538	2	2	50
TN MINOCQUA BO-DI-LAC LANDFILL	1561	2	2	50
CTY HAYWARD	1751	2	2	50
TN LINCOLN	1779	2	2	50
JUNKER SANITARY LANDFILL, INC	1972	2	2	50

SITE LIST BY PREDICTOR GROUPS

FACILITY NAME	LICENSE	# WELLS PREDICTED	TOTAL # WELLS	% TOTAL DIRTY
MAZO LAND DISPOSAL	2009	2	2	50
TN RUTLAND	2115	2	2	50
WI ST DEPT TRANSPORTATION	2586	2	2	50
CNTY SHAWANO-ANGELICA SITE	2728	2	2	50
TN ROSENDALE	2747	2	2	50
NIAGARA OF WISC PAPER CORP	3005	2	2	50
TN MENASHA	671	4	4	50
TN WESCOTT	1004	4	4	50
CTY RICHLAND CENTER	1519	4	4	50
WASTE MGMT OF WIS-MUSKEGO LF.O	2895	4	4	50
BAAP DETERRENT BURNING GROUND	3037	4	4	50
CTY MENOMONIE	372	5	5	60
RUEF'S SANITARY SERVICE, INC	478	5	5	60
GREEN LAKE SANITARY LANDFILL	1890	5	5	60
CTY JANESVILLE-ASH BEDS	3061	5	5	60
VIL WONEWOC	164	4	4	75
TN ASHWAUBENON	263	4	4	75
JONGETJES LANDFILL	943	4	4	75
LAND AND GAS RECLAMATION, INC	1118	4	4	75
WAUPACA FOUNDRY COMPANY	2638	4	4	75
CTY STOUGHTON	133	5	5	80
PELISHEK CONTRACTING LANDFILL	338	1	1	100
CTY PARK FALLS	777	1	1	100
MURRAY MACHINERY, INC	1722	1	1	100
CTY RICE LAKE	108	2	2	100
BELOIT CONCRETE STONE CO INC	781	2	2	100
SOUTHEASTERN BARRON CNTY	1887	2	2	100
CTY RIPON-TN RIPON	467	3	3	100
CTY WATERTOWN	893	3	3	100
SLINGER FOUNDRY LANDFILL	2702	3	3	100
CTY FOX LAKE WOODBURNING SITE	369	4	4	100
JAMES RIVER NORWALK-ALPINE	1832	4	4	100
REFUSE HIDEAWAY LANDFILL	1953	4	4	100
CTY WAUPUN	2246	4	4	100
CTY SUPERIOR-MOCASSIN MIKE	2627	4	4	100
CTY KENOSHA	38	5	5	100
WIS PUBLIC SERV-FLY ASH SITE	51	5	5	100
CTY LA CROSSE	144	5	5	100
CTY GREEN BAY-MILITARY AVENUE	169	5	5	100
CTY BURLINGTON	186	5	5	100
CTY TWO RIVERS	318	5	5	100
WASTE MNGT OF WIS, INC-RECLAM	1356	5	5	100
CTY MADISON-GREENTREE HILLS	1714	5	5	100
RUEF SANITARY LANDFILL	2936	5	5	100
NEKOOSA PAPERS, INC	2811	6	6	0
CNTY MONROE-RIDGEVILLE SITE	2858	6	6	0
CTY WISCONSIN DELLS	2712	8	8	0
WIS PUB SERV CORP-WESTON #3 LF	2879	8	8	0
WARD PAPER COMPANY LANDFILL	2991	7	8	0
WEYERHAEUSER COMPANY	2873	9	9	11.11
CNTY MARATHON LANDFILL	2892	7	7	14.29
WASTE CONTROL INC	1970	6	6	16.67

SITE LIST BY PREDICTOR GROUPS

FACILITY NAME	LICENSE	# WELLS PREDICTED	TOTAL # WELLS	% TOTAL DIRTY
GREDE-REEDSBURG FOUNDRY SW LF	2974	6	6	16.67
WASTE MGMT OF WIS-GREIDANUS LF	140	10	10	20
BAAP-PROPELLANT BURNING GRNDS	2814	7	7	28.57
WAUSAU PAPER MILLS LANDFILL	2875	7	7	28.57
OCEAN SPRAY CRANBERRIES, INC	2423	10	10	30
CTY ALGOMA	179	6	6	33.33
CTY PORTAGE	1885	6	7	33.33
CTY GREEN BAY-HUMBOLT ROAD	1129	6	6	50
R L O'KEEFE & SONS, INC LF	2031	6	6	50
SCOTT PAPER CO	2368	6	6	50
NEENAH-WHITING MILL LANDFILL	2576	6	6	50
SHAWANO PAPER MILLS LANDFILL	2719	6	6	50
BADGER ARMY AMMUNITION PLANT	2813	6	6	50
WIS POWER & LIGHT CO-EDGEWATER	2524	6	8	50
CTY BLACK RIVER FALLS	287	10	10	50
CTY CLINTONVILLE	314	10	10	50
CNTY GREEN	217	9	9	55.56
TN ONALASKA	507	9	9	55.56
CTY JANESVILLE	62	7	7	57.14
ANDERSON PEAT-ORGANIC COMPOS	420	7	7	57.14
JACKSON COUNTY IRON COMPANY	2924	7	7	57.14
KIMBERLY-CLARK LAKEVIEW MILL	3004	7	7	57.14
WASTE MANAGEMENT OF WIS-NEOSH	443	8	8	62.5
FALK CORPORATION	1882	8	8	62.5
CTY GREEN BAY-2130 DANZ AVENUE	170	6	6	66.67
NORTH WOODS DISPOSAL	2001	6	6	66.67
RICHLAND CENTER FOUNDRY CO	2487	6	6	66.67
NEKOOSA PAPERS,SLUDGE SPREAD	2672	6	6	66.67
BRILLION IRON WORKS, INC	2866	6	6	66.67
CTY JANESVILLE	2822	10	10	70
CTY CHIPPEWA FALLS	85	7	7	71.43
NEKOOSA PAPERS ASH-BARK SITE	1365	7	7	71.43
WI DOT HWY 100-RYAN RD	2988	7	7	71.43
CONSOLIDATED PAPERS-BIRON DIV	1687	8	8	75
CENTRAL SANITARY LANDFILL	2132	8	8	75
GENERAL MOTORS-WHEELER PIT	2795	8	8	75
CTY WHITEWATER	65	9	9	77.78
TOMAHAWK TISSUE CORP LANDFILL	1878	10	10	80
CNTY DOOR SANITARY LANDFILL	2937	10	10	80
CNTY GREEN S/W DISPOSAL SITE	2990	10	10	80
U S ARMY-BAAP ACID SPILL AREA	2934	6	6	83.33
CTY MADISON-SYCAMORE SITE	1935	7	7	85.71
VAN HANDEL SANITARY LANDFILL	49	10	10	90
WASTE MNGT OF WIS, INC-MUSKEGO	141	10	10	90
DAIRYLAND POWER COOP-GENOA #3	1747	6	6	100
MASTER DISPOSAL, INC LANDFILL	2425	7	7	100
FORT HOWARD STEEL & WIRE DIV	2972	7	7	100
U S ARMY-BAAP PERIMETER WELLS	3038	7	7	100
CNTY EAU CLAIRE-SEVENMILE CRK	2821	11	11	0
CNTY SAUK SANITARY LANDFILL	2978	11	11	0
EXXON MINERALS COMPANY	2977	13	13	0

SITE LIST BY PREDICTOR GROUPS

FACILITY NAME	LICENSE	# WELLS PREDICTED	TOTAL # WELLS	% TOTAL DIRTY
CNTY PORTAGE LANDFILL	2966	20	20	5
CTY SHAWANO - PHASE 2	3069	12	12	8.33
CNTY LA CROSSE	2637	11	11	9.09
WIS ELECTRIC POWER CO-HWY 32	2801	11	11	9.09
WIS POWER & LIGHT-NELSON DEWEY	2525	17	17	17.65
KOHLER COMPANY LANDFILL	1508	16	16	18.75
CNTY SAUK	2051	15	15	20
POPE & TALBOT WI-ABSORBENT PRD	2695	13	13	23.08
CTY ANTIGO	1357	15	15	26.67
MADISON PRAIRIE DEMOLITION LF	2918	11	11	27.27
CTY MERRILL	912	17	17	29.41
TORK LANDFILL CORP (SENECA)	2967	17	17	29.41
THILMANY PULP & PAPER CO	493	16	16	31.25
BADGER DISPOSAL	234	12	12	33.33
DAIRYLAND POWER COOP-ALMA	1673	12	12	33.33
H & R PAPER & REFUSE SERVICE	850	12	12	33.33
ROCK COUNTY-CTY JANESVILLE LF	3023	15	15	33.33
WIS POWER & LIGHT CO-COLUMBIA	2325	6	15	33.33
CTY-TN CEDARBURG	271	11	11	36.36
WIS ELEC POWR-PLEASANT PRAIRIE	2786	11	11	36.36
TN EAST TROY	24	20	20	40
WASTE MGMT OF WIS-PHEASANT RUN	1739	16	17	43.75
WASTE MGMT OF WIS-RIDGE VIEW	2575	16	16	43.75
WIS ELECTRIC POWER-OAK CREEK	2357	17	17	52.94
CTY OCONTO	137	14	14	57.14
METROPOLITAN REFUSE DIST, INC	107	14	14	57.14
CTY WEST BEND	224	18	18	61.11
DAIRYLAND POWER-OFFSITE DISP.	2927	13	13	61.54
CTY MEDFORD	341	11	11	63.64
MIDWEST DISPOSAL	73	15	15	66.67
WP&L EDGEWATER GEN STA-DRY ASI	2853	11	11	72.73
HOLTZ & KRAUSE, INC	674	12	12	75
CTY SHAWANO	2342	17	17	76.47
FLAMBEAU PAPER CORP	2756	13	13	76.92
NEKOOSA PAPERS (LIME SLUDGE)	2614	14	14	85.71
WASTE MNGT OF WIS-BROOKFIELD	1	14	14	85.71
APPLETON PAPERS, INC	30	18	18	88.89
CNTY FOND DU LAC	2358	18	18	88.89
CNTY MILWAUKEE HWY DEPT	881	11	11	90.91
TORK ALUM MUD DISPOSAL SITE	1892	12	12	91.67
WASTE MNGT OF WI, INC-POLK	307	14	14	92.86
CNTY KEWAUNEE SW BALEFILL	2975	11	11	100
CONSOLIDATED PAPERS-KRAFT DIV	1838	16	16	100
CTY SUPERIOR-WIS POINT LF	12	12	12	100
JAMES RIVER NORWALK-NORTHLAND	2893	11	11	100
NEKOOSA MILL REFUSE DISP SITE	2857	8	8	100
WASTE MANAGEMENT OF GREEN BAY	3	11	11	100
DAIRYLAND POWER COOP-CASSVILLI	96	25	25	28
CONSOLIDATED PAPERS-STEVENS P	2344	24	24	29.17
WIS ELECTRIC POWER-CEDAR SAUK	603	30	37	33.33
CNTY OUTAGAMIE	2484	33	33	42.42

SITE LIST BY PREDICTOR GROUPS

FACILITY NAME	LICENSE	# WELLS PREDICTED	TOTAL # WELLS	% TOTAL DIRTY
MOSINEE PAPER CORP. LANDFILL	2806	22	22	45.45
RHINELANDER PAPER COMPANY	1857	22	23	45.45
CONSOLIDATED PAPER WIS RIV DIV	1686	22	27	45.45
CNTY BROWN-EAST	2569	35	35	48.57
OWENS-ILLINOIS, INC LANDFILL	1346	32	40	56.25
CNTY DANE LANDFILL #2-RODEFELD	3018	32	32	59.38
LAND RECLAMATION, LTD	572	28	28	60.71
FORT HOWARD PAPER CO-GREEN BA	2332	24	24	62.5
WASTE MGMT OF WIS-OMEGA HILLS	1678	109	109	63.3
WASTE MGMT OF WIS-METRO LF	1099	42	42	64.29
NEKOOSA PAPER-WW TREATMENT R	2613	38	38	65.79
CNTY BROWN-WEST	2568	35	35	71.43
CNTY DANE LANDFILL #1-VERONA	2680	22	22	72.73
SANITARY TRANS & LF-DELAFIELD	719	51	52	76.47
CNTY WINNEBAGO	611	78	78	76.92
CONSOLIDATED PAPERS-WQC	2488	33	33	87.88
BARRETT LANDFILL, INC	1940	27	27	92.59
VULCAN MATERIALS CO	2998	15	23	93.33
TORK LANDFILL CORPORATION	652	16	29	93.75
CTY ASHLAND	177	21	21	95.24
BERGSTROM PAPER LF-NEENAH	2446	30	30	96.67
WASTE MNGT OF WIS, INC-CITY DS	37	12	12	100

APPENDIX D

Statistical Description of Clean Well Data

APPENDIX D
SUMMARY STATISTICS FOR CLEAN WELL DISTRIBUTIONS

1. Distribution of Median and IQR for background wells

2. Correlation coefficients Median vs. IQR and LOG10(median) vs LOG10 (IQR)

3. Distribution of Mean and Standard Deviation for raw datasets

4. Correlation coefficients for Mean vs. Standard Deviation and LOG10(mean) vs. LOG10(standard deviation)

5. Distribution of Lognormal Mean and Standard Deviation

6. Correlation coefficients of Lognormal Mean vs. Lognormal Standard Deviation.

7. Distribution of standardized skewness coefficient for raw datasets

8. Distribution of standardized skewness coefficient for log-transformed datasets.

SUMMARY STATISTICS FOR DISTRIBUTION OF MEDIAN AND IQR

X_1: Chloride - MEDIAN

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
12.656	34.671	2.84	1202.057	273.939	149
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
1	330	329	1885.8	201771.857	11
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
0	1	1.784	6	14.962	20.3
# > 90th %:	**Kurtosis:**	**Skewness:**			
15	65.399	7.961			

X_2: COD - MEDIAN

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
13.047	14.359	1.135	206.174	110.05	160
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
2	110	108	2087.595	60019.505	0
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
16	3.85	5	8.525	14	28.5
# > 90th %:	**Kurtosis:**	**Skewness:**			
16	16.512	3.49			

X_3: Alkalinity - MEDIAN

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
201.954	94.842	8.318	8994.993	46.962	130
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
7.7	610	602.3	26254.05	6462470.552	30
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
13	81	137.5	196.75	285	304
# > 90th %:	**Kurtosis:**	**Skewness:**			
13	1.41	.472			

GROUP I AND II WELLS ONLY

SUMMARY STATISTICS FOR DISTRIBUTION OF MEDIAN AND IQR

X_4: Conductivity - MEDIAN

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
426.646	232.326	18.425	53975.208	54.454	159
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
46	1768	1722	67836.765	3.747E7	1
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
16	177.4	271.625	382.5	573.5	695.6
# > 90th %:	**Kurtosis:**	**Skewness:**			
16	7.627	1.863			

X_5: Hardness - MEDIAN

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
231.329	121.34	10.367	14723.441	52.453	137
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
24	710	686	31692.1	9333696.06	23
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
14	87.2	137.25	211	331.25	377.8
# > 90th %:	**Kurtosis:**	**Skewness:**			
14	1.303	.781			

X_6: Iron - MEDIAN

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
.136	.234	.02	.055	171.926	137
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
.01	2.325	2.315	18.679	10.02	23
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
5	.02	.05	.08	.135	.254
# > 90th %:	**Kurtosis:**	**Skewness:**			
14	55.666	6.699			

GROUP I AND II WELLS ONLY

SUMMARY STATISTICS FOR DISTRIBUTION OF MEDIAN AND IQR

X_7: Chloride - IQR

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
7.399	18.906	1.549	357.438	255.515	149
Minimum:	Maximum:	Range:	Sum:	Sum Squared:	# Missing:
.1	173	172.9	1102.48	61058.232	11
# < 10th %:	10th %:	25th %:	50th %:	75th %:	90th %:
15	.886	1.5	3	7.1	15
# > 90th %:	Kurtosis:	Skewness:			
12	57.91	7.376			

X_8: COD - IQR

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
16.974	31.79	2.513	1010.625	187.293	160
Minimum:	Maximum:	Range:	Sum:	Sum Squared:	# Missing:
0	370	370	2715.77	206785.617	0
# < 10th %:	10th %:	25th %:	50th %:	75th %:	90th %:
15	3.3	5.275	8.84	17.275	31.775
# > 90th %:	Kurtosis:	Skewness:			
16	93.514	8.797			

X_9: Alkalinity - IQR

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
50.924	54.158	4.75	2933.113	106.351	130
Minimum:	Maximum:	Range:	Sum:	Sum Squared:	# Missing:
1.3	247.5	246.2	6620.1	715492.545	30
# < 10th %:	10th %:	25th %:	50th %:	75th %:	90th %:
13	8.25	13	29.75	60	141.5
# > 90th %:	Kurtosis:	Skewness:			
13	1.75	1.588			

GROUP I AND II WELLS ONLY

SUMMARY STATISTICS FOR DISTRIBUTION OF MEDIAN AND IQR

X_{10}: Conductivity - IQR

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
98.573	87.359	6.928	7631.508	88.623	159
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
10	795.5	785.5	15673.18	2750737.844	1
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
16	28.2	45	83	124.875	187.8
# > 90th %:	**Kurtosis:**	**Skewness:**			
16	25.869	3.955			

X_{11}: Hardness - IQR

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
51.363	44.79	3.827	2006.147	87.203	137
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
4	272	268	7036.7	634260.43	23
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
14	9.42	22.375	46	60	109.2
# > 90th %:	**Kurtosis:**	**Skewness:**			
14	7.375	2.316			

X_{12}: Iron - IQR

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
.324	.624	.053	.39	192.89	137
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
.005	4.11	4.105	44.332	67.33	23
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
12	.03	.054	.11	.272	.724
# > 90th %:	**Kurtosis:**	**Skewness:**			
14	18.639	4.057			

GROUP I AND II WELLS ONLY

SUMMARY STATISTICS FOR DISTRIBUTION OF MEDIAN AND IQR

Corr. Coeff. X_1: Chloride - MEDIAN Y_1: Chloride - IQR

Count:	Covariance:	Correlation:	R-squared:
149	630.995	.963	.927

Note: 11 cases deleted with missing values.

Corr. Coeff. X_2: COD - MEDIAN Y_2: COD - IQR

Count:	Covariance:	Correlation:	R-squared:
160	346.986	.76	.578

Corr. Coeff. X_3: Alkalinity - MEDIAN Y_3: Alkalinity - IQR

Count:	Covariance:	Correlation:	R-squared:
130	791.377	.154	.024

Note: 30 cases deleted with missing values.

Corr. Coeff. X_4: Conductivity - MEDIAN Y_4: Conductivity - IQR

Count:	Covariance:	Correlation:	R-squared:
159	13410.561	.661	.437

Note: 1 case deleted with missing values.

Corr. Coeff. X_5: Hardness - MEDIAN Y_5: Hardness - IQR

Count:	Covariance:	Correlation:	R-squared:
137	2602.428	.479	.229

Note: 23 cases deleted with missing values.

GROUP I AND II WELLS ONLY

SUMMARY STATISTICS FOR DISTRIBUTION OF MEDIAN AND IQR

Corr. Coeff. X_6: Iron - MEDIAN Y_6: Iron - IQR

Count:	Covariance:	Correlation:	R-squared:
137	.11	.753	.566

Note: 23 cases deleted with missing values.

GROUP I AND II WELLS ONLY

SUMMARY STATISTICS FOR (LOG)DISTRIBUTION OF MEDIAN AND IQR

Corr. Coeff. X_1: Chloride - LOG(MED) Y_1: Chloride - LOG(IQR)

Count:	Covariance:	Correlation:	R-squared:
149	.21	.784	.615

Note: 11 cases deleted with missing values.

Corr. Coeff. X_2: COD - LOG(MED) Y_2: COD - LOG(IQR)

Count:	Covariance:	Correlation:	R-squared:
158	.114	.805	.648

Note: 2 cases deleted with missing values.

Corr. Coeff. X_3: Alkalinity - LOG(MED) Y_3: Alkalinity - LOG(IQR)

Count:	Covariance:	Correlation:	R-squared:
130	.047	.348	.121

Note: 30 cases deleted with missing values.

Corr. Coeff. X_4: Conductivity - LOG(MED) Y_4: Conductivity - LOG(IQ...

Count:	Covariance:	Correlation:	R-squared:
159	.052	.632	.4

Note: 1 case deleted with missing values.

Corr. Coeff. X_5: Hardness - LOG(MED) Y_5: Hardness - LOG(IQR)

Count:	Covariance:	Correlation:	R-squared:
137	.042	.415	.172

Note: 23 cases deleted with missing values.

GROUP I AND II WELLS ONLY

SUMMARY STATISTICS FOR (LOG)DISTRIBUTION OF MEDIAN AND IQR

Corr. Coeff. X_6: Iron - LOG(MED) Y_6: Iron - LOG(IQR)

Count:	Covariance:	Correlation:	R-squared:
137	.191	.804	.646

Note: 23 cases deleted with missing values.

GROUP I AND II WELLS ONLY

SUMMARY STATISTICS FOR DISTRIBUTIONS OF NORMAL MEAN AND STDEV

X_1: Chloride - mean

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
14.625	35.367	2.897	1250.792	241.822	149
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
1.084	330.933	329.849	2179.129	216986.979	11
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
15	1.656	2.376	8.735	16.954	23.496
# > 90th %:	**Kurtosis:**	**Skewness:**			
15	61.749	7.68			

X_2: COD - mean

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
23.85	30.787	2.434	947.817	129.087	160
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
3.35	301.8	298.45	3815.926	241711.041	0
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
16	5.485	7.883	14.368	30.105	49.856
# > 90th %:	**Kurtosis:**	**Skewness:**			
16	40.712	5.246			

X_3: Alkalinity - mean

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
208.159	94.023	8.246	8840.403	45.169	130
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
8.232	590.8	582.568	27060.678	6773337.28	30
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
13	75.863	141.846	215.866	286.067	306.698
# > 90th %:	**Kurtosis:**	**Skewness:**			
13	.907	.253			

GROUP I AND II WELLS ONLY

SUMMARY STATISTICS FOR DISTRIBUTIONS OF NORMAL MEAN AND STDEV

X_4: Conductivity - mean

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
439.671	241.406	19.145	58276.85	54.906	159
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
47.791	1858.667	1810.876	69907.763	3.994E7	1
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
16	191.9	273.822	398.667	575.615	693.169
# > 90th %:	**Kurtosis:**	**Skewness:**			
16	9.383	2.155			

X_5: Hardness - mean

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
237.028	120.66	10.309	14558.949	50.905	137
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
25	759.733	734.733	32472.889	9677013.568	23
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
14	93.504	146.091	216.858	329.93	373.001
# > 90th %:	**Kurtosis:**	**Skewness:**			
14	2.484	.953			

X_6: Iron - mean

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
1.008	4.038	.345	16.307	400.711	137
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
.027	42.284	42.257	138.063	2356.891	23
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
14	.046	.082	.185	.413	1.581
# > 90th %:	**Kurtosis:**	**Skewness:**			
14	79.819	8.467			

GROUP I AND II WELLS ONLY

SUMMARY STATISTICS FOR DISTRIBUTIONS OF NORMAL MEAN AND STDEV

X_7: Chloride - stdev

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
9.98	26.936	2.207	725.547	269.899	149
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
.489	249.148	248.659	1487.023	122221.479	11
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
15	1.05	1.518	3.879	8.219	15.571
# > 90th %:	**Kurtosis:**	**Skewness:**			
15	49.586	6.681			

X_8: COD - stdev

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
34.305	55.907	4.42	3125.61	162.971	160
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
.956	382.388	381.432	5488.797	685265.083	0
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
16	3.986	5.768	13.146	34.946	91.681
# > 90th %:	**Kurtosis:**	**Skewness:**			
16	15.814	3.602			

X_9: Alkalinity - stdev

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
44.192	38.419	3.37	1475.989	86.936	130
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
2.399	190.974	188.575	5744.959	444283.71	30
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
13	7.381	14.917	28.262	65.508	104.697
# > 90th %:	**Kurtosis:**	**Skewness:**			
13	.816	1.182			

SUMMARY STATISTICS FOR DISTRIBUTIONS OF NORMAL MEAN AND STDEV

X_{10}: Conductivity - stdev

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
107.147	107.581	8.532	11573.598	100.405	159
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
7.329	797.197	789.868	17036.363	3654022.55	1
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
16	30.9	40.629	81.58	129.343	195.238
# > 90th %:	**Kurtosis:**	**Skewness:**			
16	14.401	3.308			

X_{11}: Hardness - stdev

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
58.964	66.683	5.697	4446.624	113.091	137
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
3.665	542.499	538.834	8078.066	1081055.801	23
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
14	8.874	23.278	40.83	68.794	126.82
# > 90th %:	**Kurtosis:**	**Skewness:**			
14	20.705	3.792			

X_{12}: Iron - stdev

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
3.104	15.595	1.332	243.218	502.448	137
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
.012	164.476	164.464	425.233	34397.528	23
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
14	.04	.077	.235	.707	3.021
# > 90th %:	**Kurtosis:**	**Skewness:**			
14	83.613	8.713			

GROUP I AND II WELLS ONLY

SUMMARY STATISTICS FOR DISTRIBUTIONS OF NORMAL MEAN AND STDEV

Corr. Coeff. X_1: Chloride - mean Y_1: Chloride - stdev

Count:	Covariance:	Correlation:	R-squared:
149	671.216	.705	.496

Note: 11 cases deleted with missing values.

Corr. Coeff. X_2: COD - mean Y_2: COD - stdev

Count:	Covariance:	Correlation:	R-squared:
160	1452.131	.844	.712

Corr. Coeff. X_3: Alkalinity - mean Y_3: Alkalinity - stdev

Count:	Covariance:	Correlation:	R-squared:
130	1245.675	.345	.119

Note: 30 cases deleted with missing values.

Corr. Coeff. X_4: Conductivity - mean Y_4: Conductivity - stdev

Count:	Covariance:	Correlation:	R-squared:
159	14886.586	.573	.329

Note: 1 case deleted with missing values.

Corr. Coeff. X_5: Hardness - mean Y_5: Hardness - stdev

Count:	Covariance:	Correlation:	R-squared:
137	3547.562	.441	.194

Note: 23 cases deleted with missing values.

GROUP I AND II WELLS ONLY

SUMMARY STATISTICS FOR DISTRIBUTIONS OF NORMAL MEAN AND STDEV

Corr. Coeff. X_6: Iron - mean Y_6: Iron - stdev

Count:	Covariance:	Correlation:	R-squared:
137	62.576	.994	.987

Note: 23 cases deleted with missing values.

GROUP I AND II WELLS ONLY

JMMARY STATISTICS FOR DISTRIBUTIONS OF log (10) of the NORMAL MEAN AND STDEV

Corr. Coeff. X_1: Chloride - log(x) of mean Y_1: Chloride - log(x) of s...

Count:	Covariance:	Correlation:	R-squared:
149	.203	.83	.689

Note: 11 cases deleted with missing values.

Corr. Coeff. X_2: COD - log(x) of mean Y_2: COD - log(x) of stdev

Count:	Covariance:	Correlation:	R-squared:
160	.184	.919	.845

Corr. Coeff. X_3: Alkalinity - log(x) of mean Y_3: Alkalinity - log(x) ...

Count:	Covariance:	Correlation:	R-squared:
130	.052	.44	.193

Note: 30 cases deleted with missing values.

Corr. Coeff. X_4: Conductivity - log(x) of mean Y_4: Conductivity - lo...

Count:	Covariance:	Correlation:	R-squared:
159	.051	.595	.354

Note: 1 case deleted with missing values.

Corr. Coeff. X_5: Hardness - log(x) of mean Y_5: Hardness - log(x) of...

Count:	Covariance:	Correlation:	R-squared:
137	.047	.443	.197

Note: 23 cases deleted with missing values.

GROUP I AND II WELLS ONLY

SUMMARY STATISTICS FOR DISTRIBUTIONS OF log (10) of the NORMAL MEAN AND STDEV

Corr. Coeff. X_6: Iron - log(x) of mean Y_6: Iron - log(x) of stdev

Count:	Covariance:	Correlation:	R-squared:
137	.461	.972	.946

Note: 23 cases deleted with missing values.

SUMMARY STATISTICS FOR DISTRIBUTIONS OF LOGNORMAL MEAN AND STDEV

X_1: Chloride - LOG MEAN

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
1.696	1.142	.092	1.304	67.335	153
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
-.281	5.669	5.95	259.515	638.455	8
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
15	.251	.625	1.842	2.684	2.959
# > 90th %:	**Kurtosis:**	**Skewness:**			
15	-.095	.292			

X_2: COD - LOG MEAN

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
2.246	.848	.067	.719	37.763	161
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
.35	4.942	4.592	361.63	927.39	0
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
16	1.008	1.79	2.132	2.851	3.312
# > 90th %:	**Kurtosis:**	**Skewness:**			
16	.265	.124			

X_3: Alkalinity - LOG MEAN

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
5.148	.673	.059	.453	13.077	131
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
2.06	6.375	4.315	674.433	3531.141	30
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
13	4.276	4.887	5.275	5.61	5.711
# > 90th %:	**Kurtosis:**	**Skewness:**			
13	5.826	-2.012			

GROUP I AND II WELLS ONLY

SUMMARY STATISTICS FOR DISTRIBUTIONS OF LOGNORMAL MEAN AND STDEV

X_4: Conductivity - LOG MEAN

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
5.911	.561	.044	.315	9.49	160
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
3.841	7.494	3.653	945.773	5640.577	1
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
16	5.184	5.586	5.966	6.32	6.529
# > 90th %:	**Kurtosis:**	**Skewness:**			
16	1.534	-.696			

X_5: Hardness - LOG MEAN

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
5.272	.597	.051	.357	11.329	138
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
3.2	6.576	3.376	727.585	3884.966	23
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
14	4.439	4.931	5.35	5.712	5.906
# > 90th %:	**Kurtosis:**	**Skewness:**			
14	.847	-.83			

X_6: Iron - LOG MEAN

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
-2.429	.906	.075	.822	-37.309	146
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
-3.997	0	3.997	-354.704	980.875	15
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
15	-3.615	-3.316	-2.465	-1.755	-1.282
# > 90th %:	**Kurtosis:**	**Skewness:**			
15	-.613	.24			

GROUP I AND II WELLS ONLY

SUMMARY STATISTICS FOR DISTRIBUTIONS OF LOGNORMAL MEAN AND STDEV

X_7: Chloride - LOG STDEV

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
.636	.327	.026	.107	51.342	153
Minimum:	Maximum:	Range:	Sum:	Sum Squared:	# Missing:
.092	1.521	1.429	97.342	78.149	8
# < 10th %:	10th %:	25th %:	50th %:	75th %:	90th %:
15	.207	.378	.587	.857	1.13
# > 90th %:	Kurtosis:	Skewness:			
15	-.312	.513			

X_8: COD - LOG STDEV

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
.971	.402	.032	.162	41.427	161
Minimum:	Maximum:	Range:	Sum:	Sum Squared:	# Missing:
.157	2.514	2.357	156.35	177.73	0
# < 10th %:	10th %:	25th %:	50th %:	75th %:	90th %:
16	.525	.69	.934	1.21	1.533
# > 90th %:	Kurtosis:	Skewness:			
16	.88	.818			

X_9: Alkalinity - LOG STDEV

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
.252	.213	.019	.045	84.54	131
Minimum:	Maximum:	Range:	Sum:	Sum Squared:	# Missing:
.026	1.28	1.254	33.046	14.249	30
# < 10th %:	10th %:	25th %:	50th %:	75th %:	90th %:
13	.044	.085	.218	.34	.537
# > 90th %:	Kurtosis:	Skewness:			
13	3.996	1.64			

GROUP I AND II WELLS ONLY

SUMMARY STATISTICS FOR DISTRIBUTIONS OF LOGNORMAL MEAN AND STDEV

X_{10}: Conductivity - LOG STDEV

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
.234	.146	.012	.021	62.434	160
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
.026	.997	.971	37.487	12.185	1
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
16	.095	.128	.191	.296	.438
# > 90th %:	**Kurtosis:**	**Skewness:**			
16	4.018	1.609			

X_{11}: Hardness - LOG STDEV

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
.269	.218	.019	.048	81.077	138
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
.017	1.37	1.353	37.17	16.545	23
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
14	.051	.119	.216	.386	.528
# > 90th %:	**Kurtosis:**	**Skewness:**			
14	5.975	1.934			

X_{12}: Iron - LOG STDEV

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
1.161	.521	.043	.271	44.891	146
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
.294	2.853	2.559	169.443	236.008	15
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
15	.508	.796	1.104	1.396	1.843
# > 90th %:	**Kurtosis:**	**Skewness:**			
15	1.201	.945			

GROUP I AND II WELLS ONLY

SUMMARY STATISTICS FOR DISTRIBUTIONS OF LOGNORMAL MEAN AND STDEV

Corr. Coeff. X_1: Chloride - LOG MEAN Y_1: Chloride - LOG STDEV

Count:	Covariance:	Correlation:	R-squared:
153	-.093	-.25	.062

Note: 8 cases deleted with missing values.

Corr. Coeff. X_2: COD - LOG MEAN Y_2: COD - LOG STDEV

Count:	Covariance:	Correlation:	R-squared:
161	-.079	-.232	.054

Corr. Coeff. X_3: Alkalinity - LOG MEAN Y_3: Alkalinity - LOG STDEV

Count:	Covariance:	Correlation:	R-squared:
131	-.049	-.339	.115

Note: 30 cases deleted with missing values.

Corr. Coeff. X_4: Conductivity - LOG MEAN Y_4: Conductivity - LOG S...

Count:	Covariance:	Correlation:	R-squared:
160	-.012	-.146	.021

Note: 1 case deleted with missing values.

Corr. Coeff. X_5: Hardness - LOG MEAN Y_5: Hardness - LOG STDEV

Count:	Covariance:	Correlation:	R-squared:
138	-.031	-.234	.055

Note: 23 cases deleted with missing values.

GROUP I AND II WELLS ONLY

SUMMARY STATISTICS FOR DISTRIBUTIONS OF LOGNORMAL MEAN AND STDEV

Corr. Coeff. X_6: Iron - LOG MEAN Y_6: Iron - LOG STDEV

Count:	Covariance:	Correlation:	R-squared:
146	.167	.354	.126

Note: 15 cases deleted with missing values.

GROUP I AND II WELLS ONLY

DISTRIBUTIONS FOR STANDARDIZED SKEWNESS COEF. -- RAW DATA --

X_1: Chloride - z-score

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
.866	.999	.115	.998	115.42	76
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
-1.454	4.462	5.916	65.786	131.807	77
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
8	-.084	.261	.71	1.244	2.045
# > 90th %:	**Kurtosis:**	**Skewness:**			
8	2.351	1.21			

X_2: COD - z-score

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
1.189	1.145	.25	1.312	96.329	21
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
.023	5.052	5.029	24.966	55.911	132
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
2	.212	.388	.897	1.532	2.433
# > 90th %:	**Kurtosis:**	**Skewness:**			
2	4.361	2.022			

X_3: pH - z-score

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
-.035	1.002	.082	1.005	-2901.311	150
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
-2.696	6.414	9.11	-5.182	149.867	3.
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
15	-.971	-.438	-.106	.228	.749
# > 90th %:	**Kurtosis:**	**Skewness:**			
15	14.713	2.633			

DISTRIBUTIONS FOR STANDARDIZED SKEWNESS COEF. -- RAW DATA --

X_4: Alkalinity - z-score

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
.143	.93	.085	.866	649.333	119
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
-2.247	4.068	6.315	17.051	104.589	34
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
12	-.882	-.325	.128	.635	.968
# > 90th %:	**Kurtosis:**	**Skewness:**			
12	2.723	.366			

X_5: Conductivity - z-score

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
.481	1.246	.101	1.552	258.765	153
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
-2.317	6.001	8.318	73.657	271.345	0
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
15	-.504	-.231	.146	.849	1.924
# > 90th %:	**Kurtosis:**	**Skewness:**			
15	4.914	1.838			

X_6: Hardness - z-score

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
.349	1.142	.1	1.303	326.807	131
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
-2.909	4.547	7.456	45.764	185.433	22
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
13	-.833	-.292	.317	.942	1.728
# > 90th %:	**Kurtosis:**	**Skewness:**			
13	1.438	.39			

GROUP I AND II WELLS ONLY

DISTRIBUTIONS FOR STANDARDIZED SKEWNESS COEF. -- RAW DATA --

X_7: Sulfate - z-score

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
.67	1.404	.222	1.971	209.621	40
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
-.933	6.191	7.124	26.789	94.806	113
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
4	-.475	-.192	.282	1.218	2.072
# > 90th %:	**Kurtosis:**	**Skewness:**			
4	5.021	2.095			

X_8: Iron - z-score

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
1.401	.882	.294	.778	62.939	9
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
.299	2.693	2.394	12.611	23.893	144
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
1	.45	.685	.997	2.302	2.627
# > 90th %:	**Kurtosis:**	**Skewness:**			
1	-1.389	.373			

DISTRIBUTIONS FOR STANDARDIZED SKEWNESS COEF. -- LOG DATA --

X_1: Chloride - log z-score

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
.055	.809	.093	.655	1473.083	76
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
-3.024	2.922	5.946	4.174	49.319	77
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
8	-.814	-.277	.047	.435	.854
# > 90th %:	**Kurtosis:**	**Skewness:**			
8	3.728	-.124			

X_2: COD - log z-score

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
.114	.435	.095	.189	382.444	21
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
-.97	.947	1.917	2.386	4.047	132
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
2	-.326	-.097	.093	.312	.726
# > 90th %:	**Kurtosis:**	**Skewness:**			
2	.53	-.224			

X_3: pH - log z-score

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
-.134	1.059	.086	1.122	-790.254	150
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
-4.208	6.389	10.597	-20.107	169.894	3
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
15	-1.025	-.521	-.174	.161	.609
# > 90th %:	**Kurtosis:**	**Skewness:**			
15	13.629	1.912			

GROUP I AND II WELLS ONLY

DISTRIBUTIONS FOR STANDARDIZED SKEWNESS COEF. -- LOG DATA --

X_4: Alkalinity - log z-score

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
-.213	1.006	.092	1.013	-472.045	119
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
-2.955	3.506	6.461	-25.368	124.897	34
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
12	-1.521	-.713	-.121	.347	.731
# > 90th %:	**Kurtosis:**	**Skewness:**			
12	2.013	-.174			

X_5: Conductivity - log z-score

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
.002	1.103	.089	1.216	50825.091	153
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
-3.56	4.758	8.318	.332	184.882	0
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
15	-.897	-.54	-.147	.482	1.32
# > 90th %:	**Kurtosis:**	**Skewness:**			
15	4.21	.89			

X_6: Hardness - log z-score

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
-.21	1.156	.101	1.337	-549.898	131
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
-4.748	2.725	7.473	-27.547	179.618	22
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
13	-1.577	-.742	-.147	.457	1.085
# > 90th %:	**Kurtosis:**	**Skewness:**			
13	2.025	-.884			

GROUP I AND II WELLS ONLY

DISTRIBUTIONS FOR STANDARDIZED SKEWNESS COEF. -- LOG DATA --

X_7: Sulfate - log z-score

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
-.22	1.055	.167	1.112	-479.55	40
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
-2.145	3.01	5.155	-8.797	45.314	113
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
4	-1.152	-.932	-.393	.319	.889
# > 90th %:	**Kurtosis:**	**Skewness:**			
4	1.955	1.119			

X_8: Iron - log z-score

Mean:	Std. Dev.:	Std. Error:	Variance:	Coef. Var.:	Count:
-.064	.642	.214	.413	-1000.238	9
Minimum:	**Maximum:**	**Range:**	**Sum:**	**Sum Squared:**	**# Missing:**
-1.66	.505	2.165	-.578	3.338	144
# < 10th %:	**10th %:**	**25th %:**	**50th %:**	**75th %:**	**90th %:**
1	-1.14	-.031	.086	.251	.436
# > 90th %:	**Kurtosis:**	**Skewness:**			
1	2.477	-1.901			

GROUP I AND II WELLS ONLY

050848- Methods for Determining Compliance with Ground-water Quality Regulations at Waste Disposal Facilities