Data Analysis Protocol for the
Ground Water Monitoring and Assessment Program

March, 1999

This material may be made available in other formats,
such as Braille, large type or audio, upon request.
TDD users call (651) 282-5332

Table of Contents

The Ground Water and Monitoring Assessment Program (GWMAP) is a non-regulatory program in the Environmental Outcomes Division of the Minnesota Pollution Control Agency (MPCA). Ground water quality data collected by GWMAP staff help

- determine ambient conditions in drinking water wells from Minnesota's major aquifers;
- identify and evaluate specific ground water issues and problems; and
- evaluate the effectiveness of pollution prevention and regulatory programs for improving or maintaining ground water quality.

This document contains a short discussion of sampling strategies (e.g. sample design, sample size), a summary of Quality Assurance/Quality Control procedures utilized in evaluating data, and step-by-step procedures for choosing a method of analyzing ground water quality data. Example analyses are included.

SPSS (Statistical Package for the Social Sciences) is the statistical software used by GWMAP[1]. Procedures described in this document are oriented toward application within SPSS. For many statistical analyses, there are suitable alternative methods which may be applied in other statistical software packages.

A short reference list accompanies this document. GWMAP has completed numerous literature reviews on a variety of environmental issues, and reference lists are available with each of those reviews. The following two documents discuss development of analysis protocols:

- Adkins, Nadine C. 1993. *A Framework for Development of Data Analysis Protocols for Ground Water Quality Monitoring*. Colorado Water Resources Research Institute. Colorado State University.
- United States Environmental Protection Agency. 1989. *Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities:Interim Final Guidance*. Office of Solid Waste, waste Management Division, United States Environmental Agency, Washington, D.C.

---

[1] Mention of a particular product does not imply preference for that product.
Abbreviations : MPCA, Minnesota Pollution Control Agency; GWMAP, Ground Water Monitoring and Assessment Program; TDS, Total Dissolved Solids; VOC, Volatile Organic Compound; UCL, Upper

The first document provides a framework for data analysis, including the role of sampling objectives. The author provides a discussion of statistical methods, including discussion of normality checks, outliers, trends (serial correlation), data censoring, the role of laboratory analysis, missing data, parametric and nonparametric descriptive statistics and hypothesis tests, and correlation analysis. A case study is included. Despite the author's attempt to keep the discussion general, the document is technical and requires knowledge of statistics on the part of the reader. While the author provides guidance on which sampling method to use to meet the study objectives, the guidance is not always clear.

The second document provides a more basic approach to understanding statistical methods in water quality analysis. The document is well written and covers relevant topics. A limitation of the document is that it is applicable only to RCRA facilities. Application to GWMAP studies requires interpretation of the methods. The discussion of trends analysis is weak. The authors recommend substituting values for censored data, and this is not recommended by GWMAP.

1. Sample Objectives, Sample Design, Laboratory Analysis

Sample objectives and design, sample collection, and laboratory analysis must be considered prior to data analysis. Sample design varies with each study. Study objectives are stated as hypotheses to be tested during the study. An appropriate and efficient sample design is then selected based on the stated hypotheses. Sample size, frequency, and analysis methods are included in the design. Sample designs are described in supporting documentation for each study conducted by GWMAP.   Montgomery (1984) provides a detailed discussion of sample design. GWMAP follows procedures described in *GWMAP Field Guidance Manual* (1996) for sample collection.

Some general guidelines followed by GWMAP are listed below:

- Factors within a design should be replicated (preferably more than twice).
- Trend studies include quarterly sampling for a period of not less than four years.

Confidence Limit; CWI, County Well Index; UTM, Universal Trans Mercator; Q_{25}, 25th quartile; Q_{75}, 75th quartile; ANOVA, Analysis of Variance.

- Individual sampling points are located to minimize spatial and serial correlation during successive sampling events.

- A minimum of 20 samples are collected from each population of interest.

- There is a preference for factorial and trend designs, both of which improve the efficiency of sample collection by reducing the number of samples required for an individual well.

Laboratories utilized by GWMAP are certified and have Quality Assurance Plans on file with the MPCA. Data censoring in the laboratory is minimized by using low reporting limits. Laboratories are expected to report measured concentration values together with statements of uncertainty such as confidence intervals. Laboratory spiked recoveries, flagged data, and blanks are examined prior to data analysis. Estimated concentrations are used for samples flagged with an "E" or "J". Samples flagged with an "E" or "J" and with no estimated concentration are assigned greater ranks for non-parametric analyses than samples flagged with a "U". Additional procedures for reviewing laboratory data are discussed in a MPCA memo (*Laboratory Data Review Check List)* dated March 6, 1997, from Luke Charpentier to MPCA staff.

2. Data Quality Analysis

Prior to statistical analysis, a series of quality checks are performed on the data. These include:

- precision estimates for field and laboratory duplicates;

- summary of results for acid blanks for cation samples;

- summary of results for spiked or standard samples;

- summary results for volatile organic compound (VOC) trip blanks;

- calculation of cation-anion balance;

- comparison of calculated and laboratory-determined total dissolved solids (TDS); and

- calculation of TDS divided by conductivity.

Precision estimates for duplicates, cation-anion balances, and precision of TDS measurements should be within 5%, although cation-anion balances and TDS comparisons of 10% or less are acceptable. The Relative Percentage Difference (RPD) is used to estimate precision of duplicates RPD = (prime - duplicate)/((prime + duplicate)/2) * 100 where prime and duplicates are concentrations of the primary and duplicate samples, respectively. Charge balances calculated using laboratory-measured bicarbonate show a consistent bias toward excess positive charge because laboratory-measured alkalinity is typically about ten percent less than field-measured alkalinity. Field alkalinity or lab-corrected alkalinity is used in calculating charge balance. Methods for calculating charge balance are described in Hounslow (1995). TDS divided by conductivity should range from 0.55 to 0.76.

If the above quality checks reveal potential problems, the following additional checks may be performed.

- Potassium concentrations should be less than sodium concentrations. Distribution of these elements is approximately equal in geologic materials, but potassium is more readily removed from solution by clay minerals and sodium commonly increases in concentration as a result of ion exchange reactions.

- The ratio of calcium to magnesium concentration reflects the rock source. Sedimentary carbonates and gypsum should have greater concentrations of calcium than magnesium, while silicate rocks have higher concentrations of magnesium.

- Calcium concentrations should be greater than or equal to sulfate concentrations because the main source of sulfate is gypsum. Greater abundance of sulfate may reflect acidic waters or calcium exchange reactions.

- Sodium concentrations should be greater than or equal to chloride concentrations, since sodium chloride is the predominant source of chloride, but there may be other sources of sodium.

GWMAP has not provided documentation of these QA/QC analyses with past reports. We anticipate having QA/QC procedures performed electronically within our database by the end of 1999. QA/QC reports will then be generated for each study.

3. Summary of Analysis Methods

Three categories of statistical analysis are utilized in GWMAP. Descriptive statistics provide information about a particular variable for a population of wells. An example is mean nitrate (the variable) concentration in water table wells (the population) across the state. The second category of analysis is hypothesis testing. These consist of comparisons between different populations of wells to determine which populations differ. The third category of analysis is correlation analysis. Statistical methods employed by GWMAP are discussed in Sections 3.1, 3.2, and 3.3. Algorithms provide information on which analysis method to use.

Figure 1 illustrates how the terms population, sample, and value are used. For example, a population consists of all wells with or all wells within a land use (i.e., residential and agriculture). Use of sub-populations is a factor analysis, with land use being the factor and the different land uses being the populations. Each well represents a sample. Nitrate concentration (in ug/L) represents a value.

Well	Land Use	Nitrate (ug/L)
1	residential	1220
2 (Samples)	residential	250 (Values)
3	agriculture	< 20
4	agriculture	10210

Populations

Figure 1 : Illustration of the terms population, sample, and value.

In Sections 3.1, 3.2, and 3.3, flowcharts guide the decision-making process. Flowcharts follow a similar format. Boxes outlined with a dash require a procedure by the user, but no analysis or decision. Solid boxes with a thin outline represent decision points in the flow chart. Solid boxes with a thick outline represent analysis points and procedures.

3.1. Descriptive Statistics

Descriptive statistics include graphical displays of data, summary statistics, and checks on population assumptions. The flow chart for conducting descriptive statistical analysis is illustrated in Figure 2.

Individual tests, analyses, and graphical methods are discussed in Section 4. Examples are provided in Section 5. **Note: choice of descriptive statistics to display are somewhat subjective. For example, users may prefer to indicate 99th percent confidence limits rather 95th percent limits, ranges instead of maximum and minimum, or percentiles instead of confidence limits.**

$\geq 1\%$ of data censored[1]

Yes   No

Single reporting limit[2]

Yes   Censor at reporting limit[3]

No

Censor at highest reporting limit[3]

Initial check of data
- Multiple observations
- Outliers and extremes
- Missing values
- Serial correlation[4]

$\geq 1\%$ of data censored[5]

No

Normal distribution
- box plots
- bar charts
- Komolgorov-Smirnov test[9]

Yes   Summary statistics[12]

Yes

Check on assumption of normal distribution[6]

log-distribution

No

Normal distribution with data transformation[10]

Yes

Conduct statistical summary using Helsel's robust method [7]

No apparent distribution   No

normal distribution

Summary statistics[11]

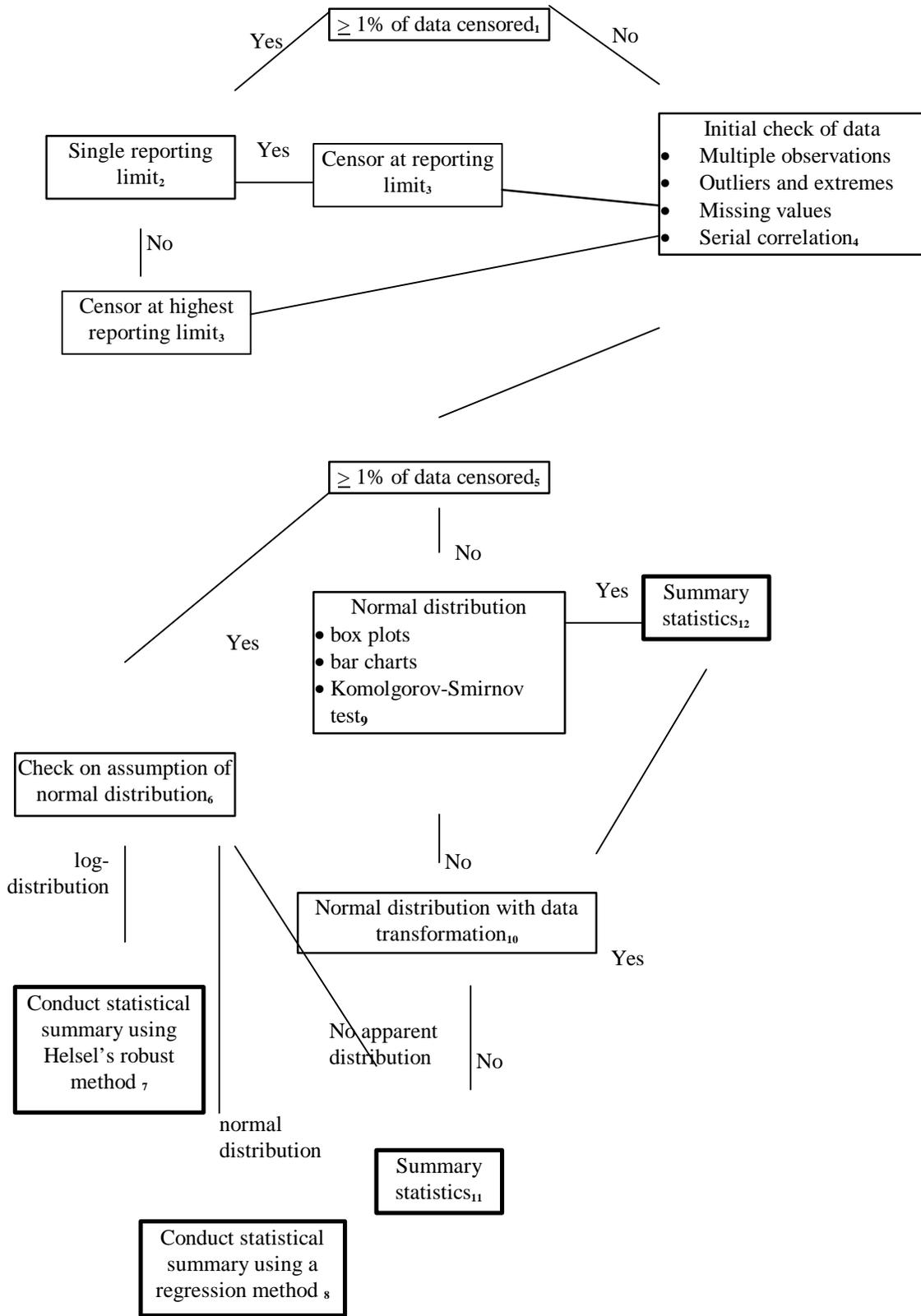Conduct statistical summary using a regression method [8]

Figure 2 : Flowchart for conducting descriptive statistical analysis of GWMAP data.

| 1 | Decision box for data censoring. If 1% or more of the values are below a reporting limit proceed to box 2. If non-detections comprise less than 1% of the values assign one-half the reporting limit to the non-detections, then proceed to box 4. |

| 2 | Decision box for multiple censoring limits. If there is more than one reporting limit, proceed to box 3, otherwise proceed to box 4. |

| 3 | Censor data at the highest reporting limit. All values less than the highest reporting limit are considered to be non-detections. For example, values of <1.0, 0.93, and <0.50 are all considered non-detections if 1.0 is the highest reporting limit in the data set. The value assigned to these data is actually below the reporting limit. For example, any value below 1.0 in the example above would serve as the censoring value, provided all censored data are assigned the same value. When completed, proceed to box 4. |

| 4 | Initial check of data. If there are multiple observations for the same sample, average them. Identify outliers but do not delete them from the analysis unless human error can be demonstrated as the source of the extreme value. Values are outliers if the value is greater than (Q_{75} + (1.5 IQR) or less than (Q_{25} - 1.5(IQR), where IQR is the interquartile range (Q_{75} - Q_{25}). A value of 3.0*IQR is used for extreme values. Missing values will only be a concern for trend analysis or for paired comparisons. Seasonality and serial correlation are considered within specific hypothesis or correlation tests. In general, seasonality is dealt with by treating season or year of sampling as an independent variable in the data analysis. Serial correlation is avoided by not using repeated sampling except within specific trend studies. |

5 Decision box for data censoring. If 1% or more of the values were censored proceed to box 7 (log-normal distribution) or box 8 (normal distribution). If non-detections comprise less than 1% of the values assign one-half the reporting limit to the non-detections, then proceed to box 9.

6 Decision box to determine if censored data represent a normal distribution. Graphically display uncensored data with a box plot and a frequency bar chart. If data appear normally distributed proceed to box 8. If data do not appear to follow a normal distribution, log-transform and display the data using a box plot and frequency bar chart. If data appear log-normally distributed, proceed to box 7. If data do not appear normally or log-normally distributed, proceed to box 11. **This box can be skipped, since the tests conducted in boxes 7 and 8 will provide output which can be used to assess the assumption of normality.**

7 Analysis : compute summary statistics using Helsel's robust method for log-normally distributed, censored data (Helsel, 1990). Summary statistics include sample size, number of non-detections, mean, confidence limits, median, maximum, and minimum. The parametric terms (mean and confidence limits) must be calculated using the slope and intercept from the Helsel regression analysis. Application examples are included in Section 5. **Note : techniques other than Helsel's method are acceptable for calculating mean and confidence limits for censored data.**

8 Analysis box for computing summary statistics using the regression method for normally distributed, censored data. Summary statistics include sample size, number of non-detections, mean, confidence limits, median, minimum, and maximum. The parametric terms (mean and confidence limits) must be calculated using the slope and intercept from the regression analysis. **Note : techniques other than the regression method are acceptable for calculating mean and confidence limits for censored data.**

9 Decision for normal distribution. Graphically display data with box plots, histograms, stem-and-leaf diagrams, and frequency bar charts. If data appear normal, test for normality using Shapiro-Wilk or Komolgorov-Smirnov tests. If the significance level of these tests is greater than or equal to 0.05, proceed to box 12. If transformation appears necessary or the significance level of tests is less than 0.05, proceed to box 10.

10 Decision box to determine if non-normally distributed data can be transformed to fit a normal distribution. Compare box plots, histograms, stem-and-leaf plots, and frequency bar charts to standard distribution curves to determine appropriate transformation. Standard distribution curves can be found in statistics text books. Transform data if a transformation seems appropriate. Display transformed data with box plots, histograms, stem-and-leaf plots, and frequency bar charts. If data appear normal, test for normality using Shapiro-Wilk or Komolgorov-Smirnov tests. If the significance level of these tests is greater than or equal to 0.05, proceed to box 12. If transformation does not seem appropriate or data cannot be transformed to a normal distribution, proceed to box 11.

11 Analysis box for censored or uncensored data which could not be fit to a normal distribution. Summary statistics for uncensored data include sample size, maximum, minimum, median, lower and upper quartiles, kurtosis, and skewness. If it is clearly indicated to potential users that data are not from a normal population, include the mean, variance, standard deviation, and coefficient of variation. For censored data, summary statistics include sample size, number of non-detections, maximum, minimum, median, lower and upper quartiles. **Note : the parametric terms must be qualified by a statement indicating the data did not fit a normal distribution, or the terms should not be reported.**
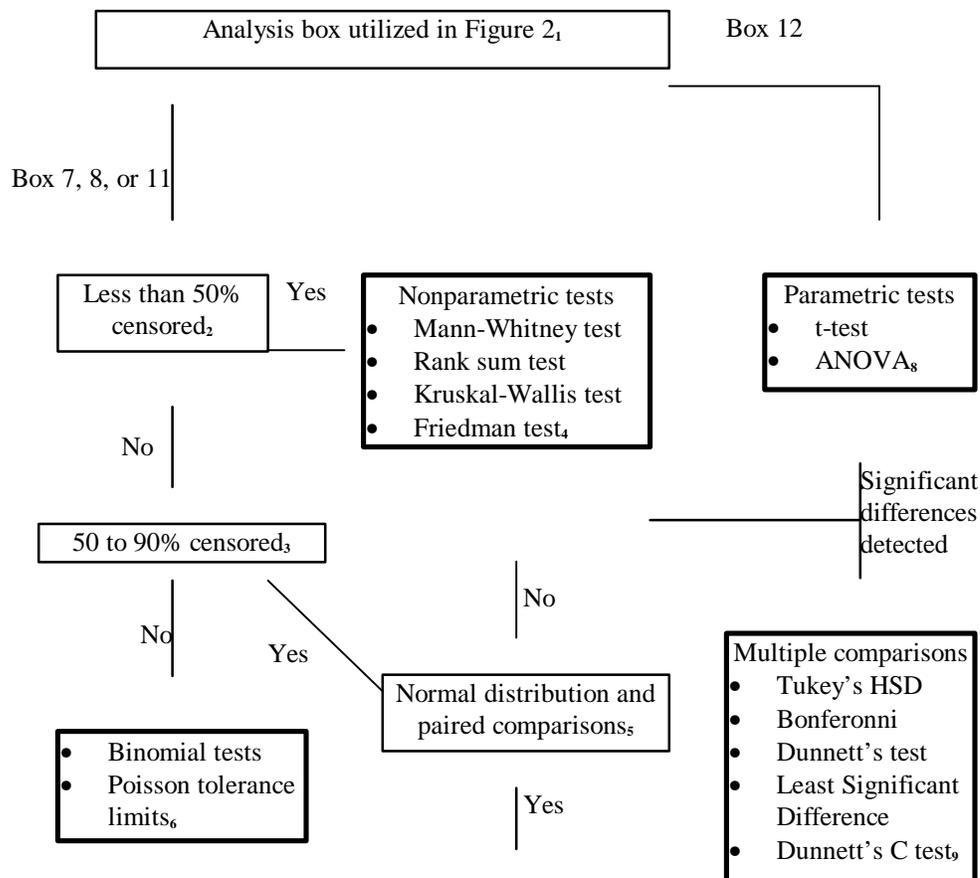
| 12 | Analysis box for uncensored data with a normal or log-normal distribution. Conduct summary statistics, including sample size, mean, variance, standard deviation, coefficient of variation, maximum, minimum, median, lower and upper quartiles, kurtosis, skewness, and confidence limits. If data have been transformed, summary analysis is performed on transformed data but summary statistics are presented in non-transformed units. |

3.2. Hypothesis Testing

Hypothesis tests compare data from different populations or to a defined value (e.g. background concentration). Hypothesis tests include t-tests, Analysis-of-Variance (ANOVA), and nonparametric equivalents of these. The flow chart for conducting hypothesis tests is illustrated in Figure 3.

Individual analysis methods are discussed in Section 4. Examples are provided in Section 5.

```
      ┌─────────────────────────────────────┐
      │   Analysis box utilized in Figure 2₁ │        Box 12
      └─────────────────────────────────────┘

Box 7, 8, or 11

┌──────────────┐  Yes   ┌─────────────────────────┐    ┌──────────────────┐
│ Less than 50%│────────│ Nonparametric tests      │    │ Parametric tests  │
│  censored₂   │        │ • Mann-Whitney test      │    │ • t-test          │
└──────────────┘        │ • Rank sum test          │    │ • ANOVA₈          │
                        │ • Kruskal-Wallis test    │    └──────────────────┘
   No                   │ • Friedman test₄         │
                        └─────────────────────────┘
                                                           Significant
                                                           differences
┌─────────────────────┐                                   detected
│ 50 to 90% censored₃ │
└─────────────────────┘                              ┌─────────────────────────┐
                                                      │ Multiple comparisons     │
   No       Yes         ┌────────────────────────┐    │ • Tukey's HSD            │
                   No   │ Normal distribution and │    │ • Bonferonni             │
                        │ paired comparisons₅     │    │ • Dunnett's test         │
┌────────────────────┐ └────────────────────────┘    │ • Least Significant       │
│ • Binomial tests    │                                │   Difference             │
│ • Poisson tolerance │          Yes                   │ • Dunnett's C test₉      │
│   limits₆           │                                └─────────────────────────┘
└────────────────────┘
```

```
┌──────────────┐
│    Tobit     │
│ regression₇  │
└──────────────┘
```

Figure 3 : Flowchart for determining analysis methods for hypothesis testing.

| 1 | Decision box for choosing parametric or nonparametric methods. If the final analysis box from Figure 2 was 8, 9, or 11, proceed to box 2, otherwise go to box 8. |

| 2 | Decision box to determine if 50% or more of the data are censored. If less than 50% are censored, proceed to box 4. If 50% or more are censored proceed to box 3. |

| 3 | Decision box to determine if 90% or more of the data are censored. If less than 90% are censored, proceed to box 5. If 90% or more are censored proceed to box 6. |

| 4 | Hypothesis tests are performed using nonparametric methods. The choice of appropriate test depends on the objectives of the analysis. Nonparametric tests are not restricted by assumptions of normal population distribution and homogeneity of variance. The choice of nonparametric method depends on the number of comparisons and ties in the data. Ties occur when more than one sample have the same value (e.g. 2 nitrate samples have a concentration of 0.50 mg/l). Non-detections are treated as ties. The procedure for choosing a nonparametric method is outlined below. |

- Comparison of one population with a fixed value (e.g. comparing a population mean to a background concentration) : one-sample chi-square test.
- Comparison between two independent samples with no ties : rank sum test (also known as the Mann-Whitney test);

- Comparison between two independent samples with ties : Wilcoxon test with ties;

- Comparison of more than two independent samples with no ties : Kruskal-Wallis test;

- Comparison of more than two independent samples with ties : Kruskal-Wallis test with ties;

- Comparison of two or more dependent samples : Friedman test.

The Pearson chi-square test is used to test the assumption of independent populations. If the significance of the Pearson chi-square test is ≥ 0.05, the populations are assumed to be independent. If the significance of the Pearson chi-square test is < 0.05, the Friedman test must be used. If the significance level of the nonparametric test is <0.05, proceed to box 9.

| 5 | Decision box to determine if data are paired and if uncensored data are from a normally distributed population. The normality assumption was assessed in box 7 of Figure 2.1.a. If data are paired and normally distributed, proceed to box 7, otherwise proceed to box 4. Most environmental data are not paired and normally distributed. If the distribution is unknown, proceed to box 4.

| 6 | Analysis box to conduct hypothesis test using Binomial tests or Poisson tolerance limits.

| 7 | Analysis box to conduct hypothesis test using Tobit regression.

| 8 | Hypothesis tests are performed using parametric methods. The choice of appropriate test is dependent on the objectives of the analysis and assumptions of sample independence and homogeneity of variance. The procedure for selecting the appropriate parametric test is outlined below.

- Comparison of one population with a fixed value (e.g. comparing a population mean to a background concentration) : one-sample t-test;

- Comparison between two independent populations with equal variance : pooled-variance t-test;

- Comparison between two independent populations with unequal variances : separate variance t-test;

- Comparison between two dependent (paired) populations : paired t-test;

- Single-factor comparison between more than two populations : One-way Analysis-of-Variance (ANOVA)

- Two-factor comparison between more than two populations : Two-way ANOVA.

Homogeneity of variance is tested using the Levene test. If the significance level of the Levene test is ≥ 0.05, the variances of the populations are assumed to be homogenous. The Pearson chi-square test is used to test the assumption of independent populations. If the significance of the Pearson chi-square test is ≥ 0.05, the populations are assumed to be independent. Methods for dealing with non-homogenous variance or dependent populations are discussed in Section 4.2.2. If the significance level of the parametric test is < 0.05, proceed to box 9.

9 If significant differences were detected from the parametric or nonparametric tests and there are more than two comparisons, compare treatment (population) means to determine which ones differ. The procedure for choosing the appropriate method is described below.

- Bonferroni test - for treatments with equal variance and when the number of treatments is four or less.

- Tukey's Honestly Significant Difference (HSD) - for treatments with equal variance and when the number of treatments is more than four.

- Dunnett's test : for treatments with equal variance which are being compared to a single control mean.

- Dunnett's C test : for treatments with unequal variance which are being compared to a single control mean.

- **Nonparametric : the test statistics discussed above (Bonferroni, Tukey, Dunnett) are applicable for determining differences from nonparametric tests, but mean ranks and the errors of the ranks are used in the analysis.**

Suitable alternatives to the above tests can be used if applicable to the data.

3.3. Correlation Testing

A correlation coefficient measures the strength of a linear relationship between two variables. Correlation tests include parametric and nonparametric tests, regression analysis, and trend analysis. The flowchart in Figure 4 illustrates the procedure for conducting correlation tests. The choice of method depends on the objectives of the study and whether a parametric or nonparametric method is utilized. Analysts need to determine which relationships are important to the overall objectives of the project.
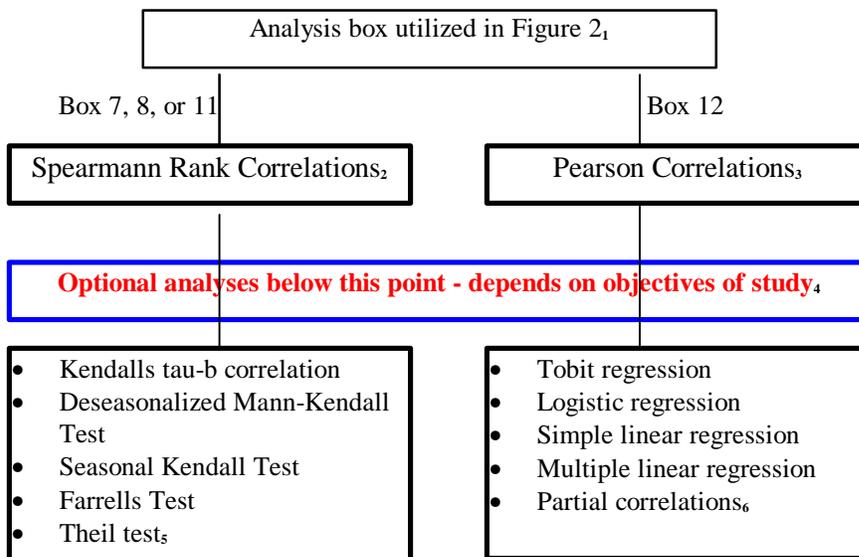

Figure 4 : Flowchart for correlation methods used in GWMAP.

[1] Decision box for choosing parametric or nonparametric methods. If the final analysis box from Figure 2 was 7, 8, or 11, proceed to box 2, otherwise go to box 3.

[2] Spearman rank correlations are performed.

3 Pearson correlations are performed.

4 Optional analyses may be performed after completing either Spearman Rank or Pearson's correlation tests. Correlations which have a significance level of less than 0.05 may be targeted for more analysis.

5 A variety of nonparametric statistical methods are available. The choice of method depends on the analysis objectives. Applicability of the methods is discussed below.

- Kendall's tau-b : an alternative to Spearmann's test; for ordered data
- Mann-Kendall test : a method for measuring monotonic trend; missing values and censored data are allowed.
- Seasoned Kendall test : a method for measuring monotonic trend, adjusted for seasonal effects; missing values and censored data are allowed.
- Theil test for slope : nonparametric method for testing linear dependence.

6 A variety of parametric statistical methods are available. The choice of method depends on the analysis objectives. Applicability of the methods is discussed below.

- Tobit regression : a form of linear regression (see below) in which the dependent and independent variables may be moderately censored. Maximum likelihood estimation (MLE) is used for the uncensored data to determine the slope and intercept of the regression. Normality checks must be conducted on the uncensored data.
- Logistic regression : predicts the probability that one or more response variables will fall into one of two discrete categories as a function of one or more independent variables. An example would be the probability that nitrate will be above a concentration of 1.0 mg/l (the two categories are yes and no) as a function of dissolved oxygen and total organic carbon concentrations.

- Simple linear regression : a method for testing the linear dependence between a dependent variable (y) and an independent variable (x), often expressed as an equation of the form $y = a + bx$, where a is the intercept and b is the slope. An example would be nitrate concentration (dependent variable) as a function of well depth (independent variable).

- Multiple regression : a method for testing the linear dependence between a dependent variable (y) and two or more independent variables $(x_1, x_2, ... x_n)$, often expressed as an equation of the form $y = a + b_1x_1 + b_2x_2 + ... + b_nx_n$, where a is the intercept and b_n is the slope for independent variable x_n. An example would be nitrate concentration (dependent variable) as a function of well depth and total organic carbon concentration (two independent variables).

- Partial correlations : Correlation tests between two or more variables, controlled for the effect of one or more variables.

4. Technical Discussion of Analysis Methods

Data analysis methods are discussed in this section. Included are equations, assumptions surrounding the analysis, and more general discussion of applicability.

4.1. Descriptive Statistics

Descriptive statistics provide information about the distribution and characteristics of a population. Means and variances (1st and 2nd statistical moments) from a normal-distributed population can be used for hypothesis testing. Means are also useful for comparison to ground water or surface water standards or limits, such as the Health Risk Limits or Aquatic Life Standards. Deviations from normality are useful in identifying cause and effect relationships. Descriptive statistics are divided into three parts in this discussion. First are graphical methods for describing a population. Second are summary statistics. Third are tests for the assumption of normal distribution.

4.1.1. Graphical Techniques

Graphical techniques provide qualitative information about the data. They are designed to observe distribution, tendencies, and extremes in the data. Box plots and histograms are the two primary graphical tools used by GWMAP. Examples are illustrated in Figures 5 and 6. Other graphical tools include stem-and-leaf diagrams and frequency bar charts.
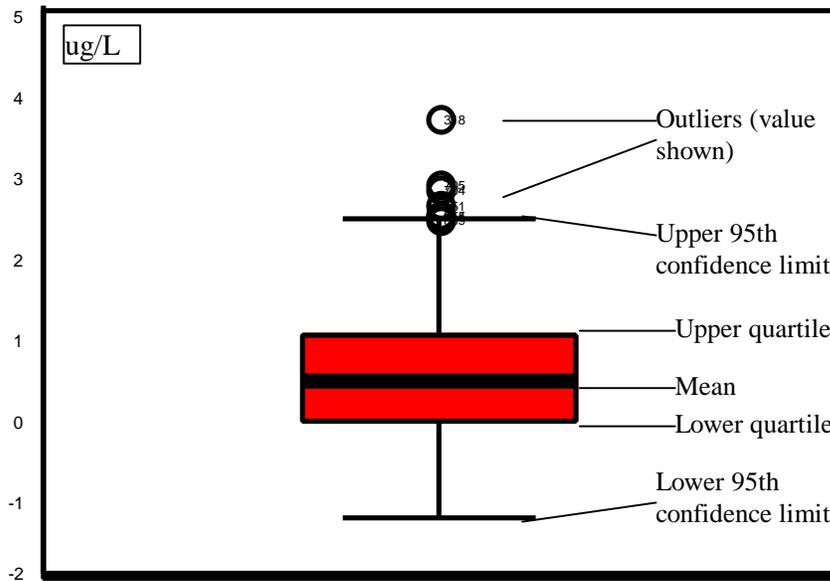
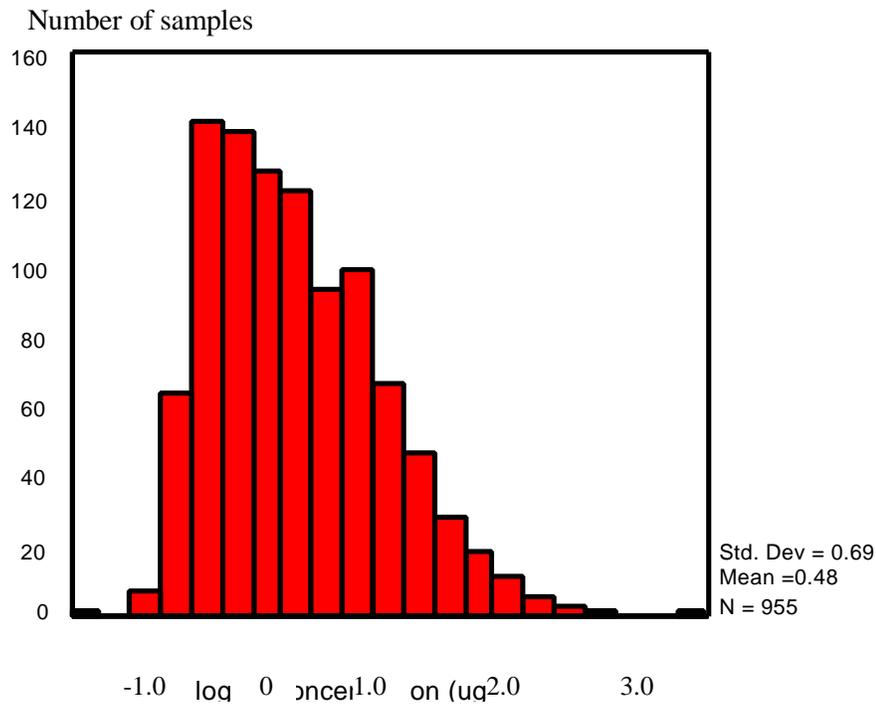

Figure 5 : Example of a box plot.

Number of samples

Figure 6 : Example of a histogram.

4.1.2. Summary Statistics

Summary statistics include the mean, variance, standard deviation, coefficient of variation, maximum, minimum, median, upper and lower quartiles, skewness, kurtosis, and 95 percent confidence limits.

The mean, μ, represents the average value and is given by

$$\mu = \sum_{i=1 \text{ to } n} (\mu_i/n)$$

where μ_i is the value of i and n is the number of samples. The median, $P_{0.5}$, represents the central value. The upper and lower quartiles (Q_{75} and Q_{25}) are the upper bounds for 75 and 25% of the values, respectively. Maximum and minimum are the largest and smallest values, respectively. Variance is a measure of spread in the data and is given by

$$\sigma^2 = \sum_{i=1 \text{ to } n} (x_i - \mu)^2/(n-1)$$

The standard deviation is the square root of the variance.

$$\sigma = \sqrt{\sigma^2}$$

The coefficient of variation is equal to the standard deviation divided by the mean

$$\sigma/\mu$$

The mean and median are approximately equal for data which conform to a standard normal distribution. In addition, about 67% of the data for a normal distribution lie within one standard deviation of the mean, about 95% within two standard deviations, and nearly 100% within three standard deviations. Skewness is a measure of the symmetry of the data and is given by

$$g_1 = \{1/n(\textstyle\sum_{i=1 \text{ to } n} (x_i - \mu)^3\}/\{1/n\sum_{i=1 \text{ to } n} (x_i - \mu)^2\}^{1.5}$$

Data which is skewed will have an asymmetrical distribution of data, with one tail being larger than the other. Values of skewness greater than 1 or less than -1 represent skewed data. Kurtosis is a measure of the peakedness of the data and is given by

$$g_2 = [\{1/n(\textstyle\sum_{i=1 \text{ to } n} (x_i - \mu)^4\}/\{1/n\sum_{i=1 \text{ to } n} (x_i - \mu)^2\}^2] - 3$$

Data with large values for the kurtosis show sharp peaks, indicating most of the data are clustered near the mean. Values of kurtosis greater than 1 or less than -1 are kurtotic.

4.1.3. Testing the Assumption of a Normal Population Distribution

Parametric statistics rely on the assumption of a normal population distribution. Slight deviations from normality typically do not have significant effects on data analysis. These small deviations include normal spread in the data resulting from factors such as sample collection procedures, variability in laboratory methods and reporting limits, and limited number of samples. For populations which are truly not normally distributed, assuming a normal distribution can result in serious errors in the analysis and incorrect conclusions.

Most environmental data are not normally distributed. More frequently, data will follow a log-normal distribution or a polynomial distribution. A polynomial distribution is one in which the data are actually from more than one population. An example would be soil samples collected from a contaminated site. Unless the entire site is either "clean" or contaminated, samples collected from the entire site will represent both "clean" and

contaminated soils, which are two separate populations. Graphical methods and an initial compilation of summary statistics can quickly reveal non-normal distributions. Signs of a non-normal distribution include:

- skewness or kurtosis greater than 3;

- mean and median differ by a factor of 2 or more;

- the coefficient of variation is greater than 100%;

- the maximum or minimum is more than five standard deviations from the mean;

- the Q25 or Q75 are more than one standard deviation from the mean.

If any of the above conditions apply, a data transformation should be considered. The most common transformation in the environmental field is the log transformation. While a log transformation often appears to provide a normal distribution, the log distribution is relatively insensitive to variability in the data and a log-transformation may mask the true distribution of data.

If none of the five signs of a non-normal distribution is observed, for transformed or untransformed data, the assumption of normality should be tested with a rigorous procedure. The Shapiro-Wilk and Komolgorov-Smirnov tests are utilized by GWMAP.

4.2. Hypothesis Testing

Hypothesis tests within GWMAP consist of comparing sample population means with other sample populations or with a constant value (e.g. a background concentration). Most comparisons are made for one factor for a single variable of interest. An example would be comparison of nitrate concentrations (variable) between different aquifer types (factor).

4.2.1. t-Tests

T-tests are used to compare means of

- two independent populations (e.g. : nitrate concentrations in wells near feedlots implementing BMPs and feedlots where BMPs are not implemented);

- paired samples (nitrate concentrations in a single well below a feedlot prior to implementation of BMPs compared to concentrations after BMPs are implemented); and

- a single variable compared to a hypothesized value (comparison of nitrate concentrations in wells below feedlots compared to background concentrations).

These three methods are illustrated schematically in Figure 7.

| Nitrate concentrations - BMPs practiced |
| Different groups of wells |
| Nitrate concentrations - BMPs not practiced |

Independent t-test : Compare means in nitrate concentrations when BMPs are practiced and not practiced

| Nitrate concentrations before BMPs practiced |
| One well sampled throughout |
| Nitrate concentrations after BMPs not practiced |

Paired t-test : Compare means in nitrate concentrations in a single well before and after BMPs are practiced

| Nitrate concentrations in wells by feedlots |

One-sample t-test : Compare mean nitrate concentrations under feedlots with background concentration
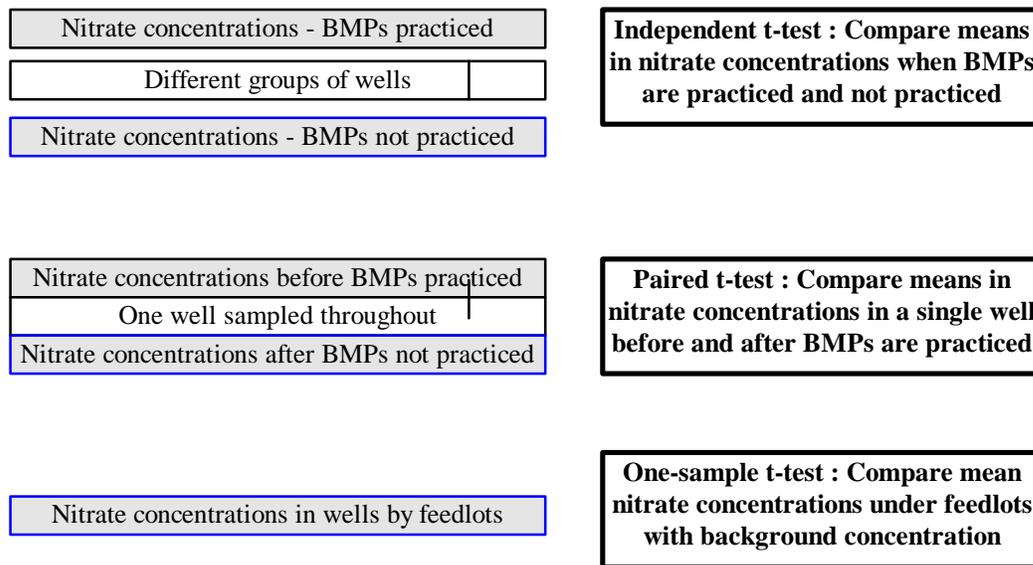
Figure 7 : Illustration of which type of t-test to use.

The test statistic utilizes the t distribution. The assumptions for a standard, pooled-variance t-test are:

- the observations are independent;
- the variances in the two groups are equal;
- the residuals of each group are normally distributed; and
- the hypothesis is (Ho : the two means are equal).

There are alternative tests for cases where observations are not independent (paired t-test) and the variances are unequal (separate variance t-test). The Levene test is used to test

the assumption of equal variance and the Pearson chi-square test is used to test the assumption of independent populations. A significance level of < 0.05 for either of these tests requires use of the appropriate alternative test. If residuals are not normally distributed and cannot be normalized through transformation, a nonparametric method should be used. The t-test is somewhat robust for slight deviations from normality. However, for the condition where there are several non-detections, there are several missing data values, or if the Komolgorov-Smirnov test for a normal population distribution test is highly significant, nonparametric methods must be utilized. Equations for t-tests may be found in the SPSS support documentation.

The test statistic for the one-sample t-test is comparison of the 95% confidence limits with the hypothesized value. Upper, lower, or two-tail confidence limits may be compared, but the t-statistic differs for one- and two-tailed comparisons.   If more than one mean is being compared to a hypothesized value, the probability calculated for each test must be adjusted by

$$p_a = pn$$

where p_a is the adjusted probability, p is the calculated probability, and n is the number of means being compared. For example, if nitrate concentrations in water-table and confined aquifer wells are being compared to background concentrations and calculated p-values were 0.03 and 0.02, adjusted p-values would be 0.06 and 0.04. For a one-sample t-test, the hypothesized value being compared to the tested value is subject to standard statistical considerations. For example, an assumed background concentration of 1.0 mg/l based on 20 observations may be less reliable than a background concentration based on 100 samples.

The t-test is preferred over the Analysis-of-Variance (ANOVA) for one-and two-sample comparisons. It is less powerful than ANOVA when more than two comparisons are being considered. For example, if means for populations A, B, and C are being compared, the ANOVA is preferred over individual t-tests comparing A and B, A and C, and B and C.

The significance level for rejecting the null hypothesis that treatment means are equal is generally 0.05. However, greater emphasis should be placed on quantifying the

probability that a particular hypothesis is true or false, compared to making decisions based on a value such as 0.05.

4.2.2. Analysis of Variance (ANOVA)

ANOVA is a method for comparing two or more groups of sample populations. One- and two-factor ANOVAs are utilized by GWMAP. The difference between one- and two-factor ANOVA is illustrated in Table 1.

Watershed	Aquifer			
	A	**B**	**C**	**D**
1	x1	x1	x1	x1
1	x2	x2	x2	x2
1	x3	x3	x3	x3
2	y1	y1	y1	y1
2	y2	y2	y2	y2
2	y3	y3	y3	y3
3	z1	z1	z1	z1
3	z2	z2	z2	z2
3	z3	z3	z3	z3

Table 1 : Example of factor designs used in ANOVA.

If x1, x2, and so on are nitrate concentrations in wells, one-way ANOVA would assess differences in mean nitrate concentrations of aquifers A, B, C, and D or watersheds 1, 2, and 3. Two-factor ANOVA would consider differences in mean nitrate concentrations between aquifer-watershed combinations (i.e. A1 vs A2 vs A3 vs B1, etc).

ANOVA utilizes a F-distribution and the general test statistic is given by F = (variation among sample means)/(variation within the samples).

ANOVA assumes:

- independent observations;

- residual follow a normal distribution;

- populations have equal variances; and

- Ho : treatment means are equal.

ANOVA is a robust analysis method and may be modified to consider transformed data, ranked data (nonparametric analysis), and unequal population variances. Additional advantages of ANOVA are that population sizes need not be equal and several pairwise comparisons of population means may be made simultaneously. If samples are not independent, additional factors may be considered to account for the lack of independence. For example, samples collected in 1994 and 1995 may not be independent because of temporal differences. Year of sampling may be included as a factor in the ANOVA, thus fulfilling the assumption of independent observations. ANOVA is fairly insensitive to deviations from the assumption of normal distribute. ANOVA also appears to be satisfactory when non-detects are replaced with one-half the reporting limit, up to about 15% of the samples being below the reporting limit. In cases where assumptions are not fulfilled, nonparametric analyses will always be more powerful.

The Levene test, Pearson chi-square test, and Komolgorov-Smirnov tests are used to test the model assumptions, as with the t-test. The significance level for rejecting the assumptions or the null hypothesis is $p < 0.05$.

If ANOVA indicates there are significant differences between treatment means, individual means can be compared to determine which ones differ. SPSS offers many options for comparing different treatments. Choice of the appropriate method is up to the discretion of the user. Additional discussion of these methods may be found in Montgomery (1984) and in the SPSS support material. Specific equations for different ANOVA are discussed in detail in Montgomery (1984). Included is a nonparametric ANOVA using the Kruskal-Wallis test.

4.2.3. Nonparametric Methods

Nonparametric methods do not have assumptions like those for the t-test and ANOVA. This is because the values (e.g. concentrations) are ranked from low to high and the statistical tests are then performed on the ranks. Ranks are insensitive to the

distribution of the data. Consequently, ties may occur, as in the case of several values below the reporting limit, but these are simply treated as equal values in the analysis. The nonparametric analysis provides information about the tendency to have larger or smaller values in one treatment compared to others.

Table 2 illustrates how values are treated in a nonparametric analysis. The data are clearly not normally distributed, with treatment A having one large value and several non-detections. When comparing means, the treatments do not differ. Comparison of ranks reveals differences between the two populations. Notice that the ranks proceed from 1 to n, where n is the total sample number. When there are ties, the average rank is used.

The Mann-Whitney test is the nonparametric equivalent of the t-test. The Kruskal-Wallis test is the nonparametric equivalent of the ANOVA. Equations for the parametric and nonparametric methods are similar and can be found in standard statistics text books. Nonparametric methods can be used for any data. If the underlying population is normally distributed, nonparametric methods will be slightly weaker than the parametric equivalents. If the underlying population is not normally distributed, nonparametric methods are much more robust than parametric equivalents. A rule of thumb may be to initially use nonparametric methods. If the resulting probability is close to some decision value, determine if parametric methods can be used. An example might be running a nonparametric test and getting a p-value of 0.06. Parametric methods, if appropriate, would provide a better estimate of the actual p-value.

Sample	Treatment A (ug/L)	Treatment B (ug/L)	Treatment A - rank	Treatment B - rank
1	< 1.0	4.3	2.5	10
2	1.1	5.1	5	14
3	1.5	5.9	6	15
4	< 1.0	4.8	2.5	12

5	1.9	4.6	7	11
6	47.7	4.9	20	13
7	< 1.0	6.9	2.5	17
8	< 1.0	8.7	2.5	18.5
9	2.8	8.7	8	18.5
10	4.1	6.1	9	16
Average	6.1	6.1	6.5	14.5

Table 2 : Comparison of data using measured and rank values.

4.3. Correlation Tests

GWMAP has utilized the Pearson test (parametric method), Spearmann rho test (nonparametric method), and least-squares linear regression. Output for the Pearson and Spearmann methods includes sample size, p-value, and correlation coefficient (R^2). The p-value defines the probability that two variables are not correlated. GWMAP uses a value of 0.05 to identify variables that are significantly correlated. The correlation coefficient defines the fraction of variability in one variable explained by variability in the other variable. The correlation coefficient is not used to identify significant correlations, but is an indicator of the strength of a correlation. For example, a correlation coefficient of 0.50 means the relationship between the two variables only explains half of the observed variability in the data.

Regression analysis is a particular form of correlation analysis in which the variability in a dependent variable (y) can be explained by the variability in a dependent variable (x). A linear regression is given by

$y = a + bx$

where a is the intercept (value of y when $x = 0$) and b is the slope (change in y over the change in x). There are four important pieces of output from a regression analysis. Two are the p-value and correlation coefficient, which are described above. The intercept and slope are the other two output items. Standard statistical packages also provide confidence limits (usually 95th percent) for the intercept and slope.

Trend analysis utilizes regression analysis, with time as the independent variable. Taylor and Loftis (1989) provide an excellent discussion of different methods for estimating trends. They recommend using the Seasonal Kendell test. A nonparametric versions of this test was developed by Hirsch and Slack (1984). GWMAP will not utilize trend analysis until four years of data have been collected for a particular study.

5. Examples

This section provides example analyses for chloride, nitrate, and Volatile Organic Compounds (VOCs). These represent a range of censoring and population distributions and are chemicals which may indicate anthropogenic impacts to ground water. Step-by-step analyses are presented for each chemical, including descriptive statistics, hypothesis testing, and correlation analysis.

The example problem will compare concentrations of chloride, nitrates, and VOCs in buried, confined Quaternary aquifers (County Well Index (CWI) code QBAA) with concentrations in water-table Quaternary aquifers (CWI code QWTA)(Wahl and Tipping, 1991). The baseline data set is the source of data for these examples (MPCA, 1998a). The baseline study was a five year investigation to determine ambient ground water quality in Minnesota's principal aquifers. We collected samples from 954 primarily domestic wells in 30 different aquifer groups. The Quaternary aquifers (QBAA and QWTA) are among the most important aquifers in Minnesota, with 506 of the 954 samples being collected from these aquifers.

The objectives of analysis were to:

- provide descriptive statistics for each aquifer;
- determine if concentrations differ between aquifers; and
- determine if concentrations are correlated with well depth, Universal Trans Mercator (UTM) easting coordinate, UTM northing coordinate, dissolved oxygen concentration, iron concentration, and redox potential.

5.1. Descriptive statistics

We utilized the flowchart in Figure 2 to calculate descriptive statistics for chloride. In Box 1, an examination of the data showed that only one value for chloride was less than the reporting limit of 200 ug/L. We proceed from Box 1 to Box 4. There were no multiple observations. Since each well was sampled just once, there was no serial correlation. There were no missing values. The Q_{25} and Q_{75} concentrations for chloride were 751 and 7170 ug/L for the QBAA aquifer and 1987 and 16986 ug/L for the QWTA aquifer, respectively. Outliers and extreme values are a distance of 1.5 and 3.0 times the interquartile range on either side of the quartiles, respectively. Outliers are thus values of 0 and 15687 for the QBAA aquifer, and 0 and 44998 for the QWTA aquifer. There were many values exceeding the outlier values for both aquifers. Proceeding to Box 5, there was less than one percent censoring. We go to box 9. Because there were so many outliers and because the lower value for outliers was 0, we proceed directly to Box 10, since chloride will not be normally distributed. Log plots indicate a non-normal distribution for the QBAA aquifer (Figure 8). We proceed to Box 11 and conduct summary statistics for the QBAA aquifer, including sample size, median, minimum, maximum, and 95th percentile. We could also display other percentiles, such as the quartiles, and parametric statistics, such as the mean. The parametric statistics would have little value for describing the distribution of chloride in the QBAA aquifer and we leave them out. Figure 9 indicates chloride may follow a log distribution for the QWTA aquifer. We compute the Komolgorov-Smirnov statistic for the QWTA aquifer. The resulting p-value of 0.200 indicates a log normal distribution adequately defines the distribution of chloride in the QWTA aquifer. We proceed to Box 12 and compute summary statistics, which include both parametric and nonparametric statistics.

Three reporting levels occurred for nitrate (20, 100, and 500 ug/L). We proceed to Box 3. The highest reporting limit is 500 ug/L. All data below a value of 500 ug/L are censored and assigned a value of 490 ug/L. We proceed to Box 4. There were no multiple observations in either aquifer and one missing value for nitrate in the QBAA aquifer. No serial correlation will be present because each well was sampled just once. Calculations of quartiles have little value because the rate of censoring was about 75

percent. We proceed to Box 5 and determine the rate of censoring to be greater than one
percent. Figures 8 and 9 illustrate log distributions for nitrate in the QBAA and QWTA
aquifers, respectively. We initially proceed to Box 7 and conduct the Helsel test. The
correlation coefficients for the regressions from this test exceeded 0.900, which suggests
the log-model was appropriate. We can now compile the summary statistics for nitrate in
each of the two aquifers. Newmann (1995) describes methods for utilizing the Helsel
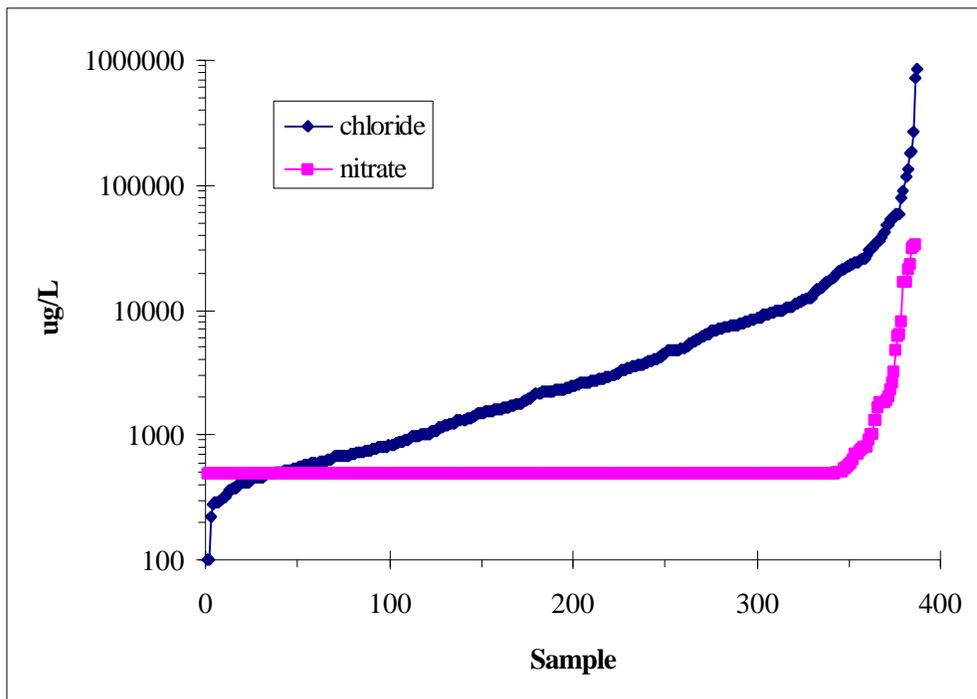method. The results are illustrated in Table 3.

Figure 8 : Distribution of chloride and nitrate data for QBAA aquifer.
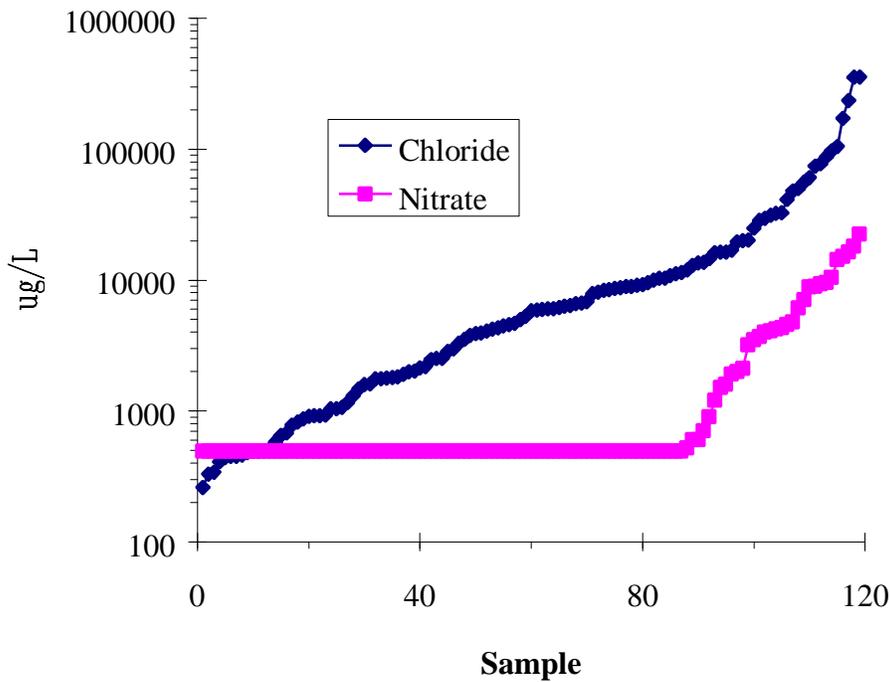
Figure 9 : Distribution of chloride and nitrate data for QWTA aquifer.

	Distribution	No. samples	No. of censored	Mean	Median	95% UCL	95th percentile	Minimum	Maximum
Chloride									
QBAA	none found	387	2	-	2320	-	39284	< 200	860510
QWTA	log-normal	119	0	5102	5810	6917	97250	260	357830
Nitrate									
QBAA	log-censored	386	342	9.0	< 500	1465	1800	< 500	33240
QWTA	log-censored	119	87	310	< 500	8348	10400	< 500	22300

Table 3 : Summary statistics for chloride and nitrate in the QBAA and QWTA aquifers.

   Although we have completed the analysis of descriptive statistics for chloride and nitrate, it is useful to review the data. For chloride in the QBAA aquifer, note that a mean and 95% UCL are not reported, since a distribution was not established. For chloride in the QWTA aquifer, the mean and median are reasonably close, as would be expected. The 95% UCL and 95th percentile are not close, however. The chloride data are skewed to higher concentrations.

The 95th percentile probably provides better information for the distribution of chloride in the high range of concentrations. This is because the sample size was large (n = 119) and the data are probably representative of the actual distribution of chloride in QWTA aquifers. For nitrate, 95% UCLs and 95th percentiles are reasonably close.

Little can be done with descriptive statistics for VOCs. The frequency of detection was very low. We are probably more interested in the frequency of detecting VOCs rather than establishing descriptive statistics for individual VOCs.

5.2. Hypothesis Testing

The flowchart in Figure 3 is utilized to test hypotheses. For this example, we will test three hypotheses.

1. Concentrations of nitrate and chloride do not differ between the QBAA and QWTA aquifers.

2. Concentrations of nitrate do not differ by well diameter in the QBAA and QWTA aquifers.

3. The frequency of VOC detection is not different between the QBAA and QWTA aquifers.

In hypothesis 1, we start in Box 1. Since we used Boxes 7 and 11 for establishing descriptive statistics we proceed to Box 2 in Figure 3. The censoring level for chloride was less than 50 percent and we proceed to Box 4. The censoring level was between 50 and 90 percent for nitrate and we proceed to Box 3 and then Box 5. The samples were not paired so we proceed to Box 4 for nitrate. Since we are comparing two treatments, we use the Mann-Whitney test. The p-value for these tests was 0.0003 and < 0.0001 for chloride and nitrate, respectively. This provides very strong evidence the hypothesis is not correct and there are significant differences between the two aquifers in concentrations of both chemicals. Mean ranks (remember the method is nonparametric) for nitrate were 243 and 285 for the QBAA and QWTA aquifers, respectively. We conclude concentrations of nitrate are greater in the QWTA aquifer compared to the QBAA aquifer.
Mean ranks for chloride were 241 and 296 for the QBAA and QWTA aquifers, respectively. We conclude concentrations of chloride are greater in the QWTA aquifer.

To conduct the second hypothesis, we establish each well diameter as a separate treatment. There were seven well diameters (4, 5, 6, 12, 24, 30, and 36 inch) for the QBAA aquifer and six well diameters (4, 5, 6, 24, 30, and 36 inch) for the QWTA aquifer, respectively. This means there were seven and six treatments for the QBAA and QWTA aquifers, respectively. From the first hypothesis, we know we will end up in Box 4. Because we have multiple comparisons, we use the Kruskal-Wallis test. The resulting p-values were < 0.001 for both aquifers, providing strong evidence that nitrate concentrations differed by well diameter. To compare nitrate concentrations in different treatments (i.e. between different well diameters), we choose a method from Box 9. The Least Significant Difference (LSD) method was chosen for this example. We need to define a significance level for this test and choose a value of 0.05. The resulting LSDs were 166 and 42 for the QBAA and QWTA aquifers, respectively. This means that any well diameters in which mean ranks differed by more than these values are considered to have nitrate concentrations which differ at the 0.05 level. Mean ranks are illustrated in Table 4. Comparing well diameters for the QBAA aquifer, a clear break is evident for well diameters exceeding 6 inches in diameter. Nitrate concentrations in wells with diameters exceeding 6 inches are significantly greater than concentrations in wells with diameters of 6 inches or less. Such a clear break in the data is rare, particularly considering the level of censoring which occurred. Well diameter appears to be an important factor affecting nitrate concentrations in the QBAA aquifer. The same general trend is apparent in the QWTA aquifer, but the break in the data is less distinct.

NOTE : Hypothesis tests are useful for identifying differences between populations, but they provide no information on physical processes causing these differences. In the case of well diameter, for example, the cause of high nitrate concentrations in large diameter wells is related to well construction, rather than the diameter of the well. Large diameter wells are often dug wells and have joints which allow infiltration of water throughout the well length.

Considering the third hypothesis in Box 3, we calculate a VOC detection frequency of 13 and 10 percent for the QWTA and QBAA aquifers, respectively. We proceed to Box 6 because the most appropriate test is the binomial test because we are interested in a

detection rather than a concentration. The resulting population can therefore be divided into detects and non-detects, creating a binomial distribution. We again need to define a level of confidence, which we assume to be 0.05. The upper 95th percent probability of detecting a VOC in QBAA wells was calculated to be 12.4 percent, which is less than the observed value of 13.4 percent in the QWTA aquifer. We conclude that the probability of detecting a VOC in the QWTA aquifer is greater than that in the QBAA aquifer.

   NOTE : Statistical analysis has again revealed significant differences between populations, but no physical explanation is provided by this analysis. The final step in the analysis process is to identify the physical factors that lead to the observed differences. In the case of these two aquifers and VOCs, QWTA aquifers are more susceptible to contamination because they are readily recharged by water percolating through the vadose zone. They are also, on average, closer to contaminant sources than QBAA aquifers.

Diameter (inches)	Mean Rank - QBAA	Mean Rank - QWTA
4	252 a	59 a
5	249 a	66 ab
6	281 a	45 a
12	457 b	-
24	491 b	104 bc
30	501 b	110 c
36	503 b	115 c
LSD	166	42

Table 4 : Mean rank nitrate concentrations for different well diameters in the QBAA and QWTA aquifers. Well diameters with different letters in a column differ in mean rank nitrate concentration.

5.3. Correlation Tests

   Hypothesis testing indicated chloride and nitrate concentrations and the frequency of detecting a VOC were greater in the QWTA aquifer than in the QBAA aquifer. We may wish to identify some factor(s) contributing to these results. Hypothesis testing is useful for identifying broad differences between or within a population, but correlation analysis is

more useful when attempting to identify cause-and-effect relationships. This is because the variables involved in the analysis are continuous, compared to hypothesis testing, where variables are divided into treatments or classes.

A good understanding of the physical system can be useful in identifying appropriate correlation tests. For example, nitrate is affected by oxidation-reduction (redox) reactions, chloride is not, and some VOCs are. We would not conduct correlation tests between chloride and redox variables, but we do for nitrate and VOCs. The following correlation tests were conducted for the QWTA aquifer.

1. Nitrate correlation with iron, manganese, dissolved oxygen concentrations, redox potential, well depth, well diameter, chloride, static water level, and UTM-east coordinate.

2. Chloride correlation with potassium, sodium, calcium, nitrate, well depth, well diameter, static water level and UTM-east coordinate.

3. Halogenated aliphatic concentration with bicarbonate (alkalinity), chloride, dissolved oxygen, iron, manganese, nitrate, and sulfate concentration, pH, redox potential, static water level, and well depth.

Because nitrate and VOCs were highly censored, we choose nonparametric methods for conducting the correlation tests. Nonparametric tests were also conducted for chloride, although parametric tests could be conducted for log-transformed data. Spearmann correlations were chosen for all the analyses. No additional correlation tests were conducted.

The correlations for nitrate are illustrated in Table 5. Iron, manganese, dissolved oxygen, and redox potential represent redox parameters, while well diameter, static water elevation, UTM-east coordinate, and well depth represent physical or hydrologic factors which may impact concentrations.

Chloride was included because it is another factor being considered in other correlation tests. The results indicate nitrate was significantly correlated ($p < 0.05$) with iron, manganese, dissolved oxygen, redox potential, well depth, chloride, and UTM-east coordinate. The strongest correlation was with iron, with a correlation coefficient of -0.654. This means 65.4% of the variability in nitrate concentration could be explained by

variability in iron concentration and that nitrate concentrations decreased as iron concentrations increased. Relatively strong relationships were observed with manganese, dissolved oxygen, redox potential, and chloride, while the correlations with well depth and UTM-east coordinate were weaker. Redox conditions within an aquifer exert a strong control on nitrate concentration.

Variable	p-value	Correlation coefficient
Iron	< 0.001	-0.654
Manganese	< 0.001	-0.464
Dissolved oxygen	< 0.001	0.397
Redox potential	< 0.001	0.544
Well depth	0.014	-0.221
Well diameter	0.121	0.143
Chloride	< 0.001	0.383
Static water level	0.827	-0.020
UTM-east	0.041	-0.185

Table 4 : Results of correlation analysis between nitrate and selected variables.

Table 5 illustrates correlations for chloride. Chloride was correlated with potassium, sodium, and calcium, but not with well depth, well diameter, and static water level. The results suggest a correlation between dissolved species and chloride. Chloride is affected by two processes. Anthropogenic sources include road salts, fertilizer, and animal wastes. Dissolution of aquifer material also results in increases in chloride concentration. These lead to increased concentrations of total dissolved solids, including chloride, but they occur under different physical conditions. Anthropogenic inputs are greatest near the top of an aquifer, while dissolution inputs are greatest deeper in the aquifer. Consequently, these two factors negate each other and there is no affect of well depth.

Variable	p-value	Correlation coefficient

Potassium	0.002	0.278
Sodium	< 0.001	0.375
Calcium	< 0.001	0.484
Nitrate	< 0.001	0.383
Well depth	0.415	-0.074
Well diameter	0.121	0.143
Static water level	0.779	-0.026
UTM-east	0.022	-0.207

Table 5 : Results of correlation analysis between chloride and selected variables.

Halogenated VOCs are persistent in ground water and are degraded under reducing conditions. VOC concentrations were positively correlated with dissolved oxygen and nitrate concentration, and redox potential, reflecting the persistence of these chemicals in oxygenated ground water. A negative correlation with pH was observed. A mechanism of chemical degradation for halogenated VOCs in ground water is through hydrolysis, in which a hydroxyl ion replaces a halogen in the chemical structure. Greater activity of the hydroxyl ion (high pH) leads to a reduction in concentration of halogenated VOCs.

Variable	p-value	Correlation coefficient
Alkalinity	0.661	0.042
Chloride	0.618	0.047
Well depth	0.263	-0.105
Dissolved oxygen	0.014	0.229
Variable	**p-value**	**Correlation coefficient**
Redox Potential	0.004	0.266
Iron	0.279	-0.102
Manganese	0.217	-0.116
Nitrate	0.010	0.238
pH	0.013	-0.230

Sulfate	0.391	0.081
Static water level	0.976	0.003

Table 6 : Results of correlation analysis between halogenated VOCs and selected variables.

6. Additional Considerations

This document represent a set of recommended guidelines for analyzing information collected by GWMAP. We have completed reports for the statewide baseline study (1998a), the St. Cloud land use study (MPCA 1998b, 1998c, 1999a), and a septic study in Baxter, Minnesota (MPCA, 1999b). We have conducted extensive data analysis for each of these reports. We have not followed these guidelines rigorously in these reports. There are several reasons for this. First, we have not had a mechanism for generating QA/QC reports. Generating these reports is time-consuming. We conduct a QA/QC analysis for each data set, but this analysis is not documented in our reports. We are developing a new database which will allow us to generate QA/QC reports for each set of analyses we conduct. Second, we often do not utilize the nonparametric methods recommended in Figure 3. We have found the Mann-Whitney and Kruskal-Wallis tests to be nearly as powerful as other methods, even when there are a large number of ties (usually non-detections) in the data. These tests are easy to run in SPSS and easy to interpret. Finally, we have not conducted studies for sufficient lengths of time to conduct trend analysis. This will be an important component of our program in the future.

Analysis procedures within GWMAP will continue to change as objectives of the program change. Examples of analyses not currently used within GWMAP are discussed below.

a) Comparison of impacted wells to a compliance well. An example would be comparison of wells near feedlots where BMPs have recently been implemented to a background well or a well near a feedlot where BMPs have been utilized for a long period of time.

b) Comparison of wells from a population to concentration limits (e.g. drinking criteria).

c) Intra-well trend analysis to determine if water quality within an individual well is changing over time.

d) Inter-well trend analysis to determine if water quality in wells from a group or groups of populations is changing over time.

We also feel it is important to clarify the utility of statistical analyses. Statistical methods allow us to design our studies so we can test appropriate hypotheses. Statistical methods then provide us with information about the validity of our design and the significance of our hypotheses. Statistical methods can be used to identify important factors affecting water quality. Statistics are not a substitute for physical interpretations. Study designs and interpretations of statistical results only have meaning if the physical processes impacting ground water are understood.

References

The following reference list represents a compilation of literature cited in this report and additional references which may be useful to people analyzing ground water data.

Frapporti, G., S.P. Vriend, and P.F.M. van Gaans. 1994. *Qualitative Time Trend Analysis of Ground Water Monitoring Networks*. Environ. Monit. and Assess. 30:81-102.

Helsel, D.R. 1990. *Less than Obvious - Statistical Treatment of Data below the Detection Limit*. Environ. Sci. Tech. 24:1767-1774

Hounslow, A.W. 1995. *Water Quality Data - Analysis and Interpretation*. Lewis Publishers. Boca Raton, Florida. 397 pp.

Hult, M.F. 1979. *Design of a Network for Monitoring Ground Water Quality in Minnesota*. U.S. Geological Survey Open file Report 79-1164. St. Paul, MN. 44pp.

Liu, S., J. Lu, D.W. Kolpin, and W.Q. Meeker. 1997. *Analaysis of Environmental Data with Censored Observations*. Environ. Sci. Tech. 31:3358-3362.

Meyer, P.D., A.J. Valocchi, and J.W. Eheart. 1994. *Monitoring Network Design to Provide Initial Detection of Ground Water Contamination*. Wat. Res. Res. 30:2647.

Minnesota Pollution Control Agency. 1996. *GWMAP Field Guidance Manual*. St. Paul, MN 42p.

Minnesota Pollution Control Agency. 1998a. *Baseline Water Quality of Minnesota's Principal Aquifers.* St. Paul, MN. 145 p. and appendices..

Minnesota Pollution Control Agency. 1998b. *Effects of Land Use on Ground Water Quality. St. Cloud Area, Minnesota. 1997 Results*. St. Paul, MN 59 p.

Minnesota Pollution Control Agency. 1998c. *Water Quality in the Upper Fifteen Feet of a Shallow Sand Aquifer in a Variable Land Use Setting - St. Cloud Area, Minnesota*. St. Paul, MN 52 p.

Minnesota Pollution Control Agency. 1999a. *Effects of Land Use on Ground Water Quality. St. Cloud Area, Minnesota. 1998 Results*. St. Paul, MN. Draft in review.

Minnesota Pollution Control Agency. 1999b. *Effects of Septic Systems on Ground Water Quality - Baxter, Minnesota*. St. Paul, MN. Draft in review.

Nelson, J.D., and R.C. Ward. 1981. *Statistical Considerations and Sampling Techniques for Ground-Water Quality Monitoring*. Ground Water. 19:617-625.

Sgambat, J.P., and J.R. Stedinger. 1981. *Confidence in Ground Water Monitoring*. Ground Water Monitoring Review. 1(1):62-69.

Snyder, W.M., A.W. Thomas, A.L. Dillard, and W.C. Mills. 1996. *Quantification for Visualization of Time Trend*. Wat. Res. Bull. 32:3:521-529.

Stoline, M.R., Passero, R.N., and Barcelona, M.J. 1993. *Statistical Trends in Ground Water Monitoring Data at a Landfill Superfund Site : A Case Study*. Environ. Monitor. Assess. 27:201-219.

Taylor, C.H., and J.C. Loftis. 1989. *Testing for Trend in Lake and Ground Water Quality Time Series*. Water Res. Bull. 25:715-726.

Tobias, F.J., J. Bech, and P.S. Algarra. 1997. *Statistical Approach to Discriminate Background and Anthropogenic Input of Trace Elements in Soils of Catalonia, Spain*. Water, Air, and Soil Poll. 100:63-78.

United States Geological Survey. *Evaluation of Regional Ground-water Quality in Relation to Land-Use.* USGS Water Resources Investigations Report 84-4217. U.S. Geological Survey Toxic Waste - Ground-Water Contamination Program.

Wahl, T.E., and R.G. Tipping. 1991. *Ground-water Data Management - The County Well Index*. Minnesota Geological Survey. Minneapolis, MN. 38 pp.

Yu, Y., and S. Zou. 1993. *Relating Trends of Principal Components to Trends of Water-Quality Constituents*. Water Res. Bull. 29:797-806.

Yu, Y., S. Zou, and D. Whittemore. 1993. *Non-parametric Trend Analysis of Water Quality Data of Rivers in Kansas*. Journal of Hydrol. 150:61-80.