**SPECIAL STUDY FOR THE
STATISTICAL EVALUATION OF
GROUNDWATER DATA TRENDS**

FINAL

MAY 1993

Work Performed Under Contract No. DE-AC04-91AL62350

This report supersedes report number UMTRA-DOE/AL-400693.0000

**Prepared for
U.S. Department of Energy
UMTRA Project Office
Albuquerque, New Mexico**

**Prepared by
Jacobs Engineering Group Inc.
Albuquerque, New Mexico**

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

RWR

DISCLAIMER

**Portions of this document may be illegible
in electronic image products. Images are
produced from the best available original
document.**

EXECUTIVE SUMMARY

Analysis of trends over time in the concentrations of chemicals in groundwater at Uranium Mill Tailings Remedial Action (UMTRA) Project sites can provide valuable information for monitoring the performance of disposal cells and the effectiveness of groundwater restoration activities. Random variation in data may obscure real trends or may produce the illusion of a trend where none exists, so statistical methods are needed to reliably detect and estimate trends.

Trend analysis includes both trend detection and estimation. Trend detection uses statistical hypothesis testing and provides a "yes" or "no" answer regarding the existence of a trend. Hypothesis tests try to reach a balance between false negative and false positive conclusions. To quantify the magnitude of a trend, estimation is required. Confidence intervals provide a range of possible values for the true slope of the concentration trend over time. Confidence intervals also involve a balance between precise and reliable estimation of the trend.

Two statistical approaches to trend analysis are parametric linear regression and nonparametric methods. Linear regression is at most 10 percent more efficient than the nonparametric procedures under certain ideal conditions based on classical normal theory. However, these ideal conditions are unlikely to occur at UMTRA Project sites. Outliers, which are occasional large errors in measurements, are present in many UMTRA Project data sets. When data are prone to outliers, nonparametric methods can be much more than 10 percent more efficient than linear regression. Therefore, nonparametric methods are recommended for trend analysis of UMTRA Project groundwater data. Linear regression maintains some advantages for fitting more complicated trend models.

Many UMTRA Project data sets contain censored values for water-quality measurements—values reported as being less than the reported laboratory detection limit. Trend analysis is complicated when censored data are present because actual measurements are unavailable for calculations. If a number is substituted in place of the censored values, calculations are possible, but the results will be affected by the choice of number used in the substitution. Nonparametric procedures are less affected by the choice than are parametric methods. A parametric alternative to substitution, called TOBIT regression, estimates trend by the well-established statistical method of maximum likelihood. An advantage to TOBIT regression is that it uses all the information available in censored values and does not require the substitution of numbers. The disadvantage of this procedure is that it relies on normal theory, which cannot be expected to be appropriate for many UMTRA Project data.

More than one detection limit is common in UMTRA Project groundwater quality data sets, adding additional difficulties to trend analysis. Trends in the detection limit can be mistaken as representing a trend in concentration if substitution methods are used. Two alternative approaches for handling data sets with multiple detection limits are possible at the present time. TOBIT regression can be used if the required model assumptions are met. The Mann-Kendall trend test is also possible, provided that all data known to be less than the highest detection limit are viewed as being equal to each other. Neither approach is very satisfactory for UMTRA Project data.

The ability of trend-analysis methods to detect and measure trends is influenced by many factors. Improvements in water sampling and laboratory analysis that reduce the variability of measurements will increase the efficiency of trend analysis. If the variability of measurements can be reduced by 25 percent, for example, then 25 percent fewer measurements will be required for equivalent trend analysis efficiency. Procedural changes that may affect the bias of measurements should be documented and examined carefully so that changes in measurement bias are not mistaken for a trend in concentration. Timing of sampling rounds at a site should be based on knowledge of seasonal cycles in water quality at that site. Lastly, sampling less frequently than twice a year may be cost effective, depending on how important it is to detect a trend quickly.

Recommendations for the UMTRA project are as follows:

- Computer software is needed to facilitate trend analysis computations and interpretation.
- Training of UMTRA Project staff is recommended to maximize the potential benefits of trend analysis for site management.
- Trends analysis should be included as part of the ongoing water sampling and analysis plan (WSAP) efforts and data validation activities.
- More investigation is needed into methods of analyzing censored data with multiple detection limits.

TABLE OF CONTENTS

<u>Subject</u>	<u>Page</u>
1.0 INTRODUCTION	1-1
2.0 STATISTICAL CONCEPTS FOR TREND ANALYSIS	2-1
2.1 Kinds of trends	2-1
2.2 Error distributions	2-1
2.2.1 The mean value of an error distribution and bias	2-3
2.2.2 Variability of error distributions	2-3
2.2.3 Shape of error distributions	2-3
2.2.4 Changes in error distributions over time	2-3
2.3 Hypothesis tests	2-4
2.3.1 The null and alternative hypotheses in trend analysis	2-4
2.3.2 H_0 and H_a , level of significance, and Type 1 errors	2-8
2.3.3 Type 2 error and power	2-9
2.4 Trend estimation	2-9
2.4.1 Sampling distribution of a trend estimator	2-9
2.4.2 Confidence intervals	2-10
3.0 STATISTICAL TREND METHODS	3-1
3.1 Linear regression	3-1
3.1.1 The simple linear regression model	3-1
3.1.2 Other linear regression models	3-2
3.1.3 Checking model assumptions	3-2
3.2 Nonparametric methods	3-3
3.3 Example	3-5
3.3.1 Parametric linear regression analysis	3-7
3.3.2 Nonparametric trend analysis	3-7
3.4 Comparison of parametric and nonparametric trend methods	3-9
3.4.1 Relative efficiency of the two methods	3-9
3.4.2 Other considerations in selecting a trend method	3-12
3.5 Recommendations for selecting a trend analysis method	3-13
4.0 TREND DETECTION WITH CENSORED CONCENTRATION MEASUREMENTS ..	4-1
4.1 Single detection limits	4-1
4.2 Multiple detection limits	4-1
5.0 EFFECT OF DATA COLLECTION METHODS ON TREND ANALYSIS	5-1
5.1 Methods of data collection and analysis	5-1
5.2 Timing of sampling	5-1
5.3 Frequency of sampling	5-3
6.0 SAMPLE SIZES NEEDED FOR TREND DETECTION	6-1
6.1 The signal-to-noise ratio	6-1
6.2 Power calculations	6-1

TABLE OF CONTENTS (Concluded)

<u>Section</u>	<u>Page</u>
7.0 CONCLUSION AND RECOMMENDATIONS	7-1
8.0 REFERENCES	8-1
9.0 LIST OF CONTRIBUTORS	9-1

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
2.1 Types of trends in concentration over time	2-2
2.2 Error distribution	2-5
2.3 Changes in error distribution over time	2-6
2.4 Error distribution imposed on a real trend in average concentration	2-7
3.1 Boxplots of nitrate measurements at six background wells, Tuba City, Arizona .	3-4
3.2 Parametric linear regression analysis of uranium data	3-6
3.3 Example of the lack of efficiency of parametric procedure when an extreme outlier is present	3-11
5.1 Effect of timing of semiannual sampling on trend analysis when concentration has seasonal variation	5-2
5.2 Effect of sampling interval on trend estimation	5-4

LIST OF TABLES

<u>Table</u>	<u>Page</u>
3.1 Uranium concentration in well 906 during 1988-91, Tuba City, Arizona	3-5
3.2 Nonparametric trend analysis of uranium data	3-8
3.3 Comparison of parametric and nonparametric analyses of uranium data	3-9
6.1 Power or probability of detecting a statistically significant trend in a water quality variable	6-2

CHANGE HISTORY

Document version	Date	Pages/Comments
Ver. 1	1/93	Initial version
Ver. 2	5/93	Version 2 represents a complete revision of Version 1; therefore, sidebar or redline indications of document changes are not used. TAC review comments incorporated.

1.0 INTRODUCTION

Monitoring for groundwater quality at Uranium Mill Tailings Remedial Action (UMTRA) Project sites has taken place since the early 1980s and will continue after surface remediation is complete and in conjunction with groundwater restoration activities. Analysis of trends over time in concentrations of chemicals in the water will be an important component of both the surface and groundwater programs in the future.

The surface program will involve continued groundwater monitoring at completed disposal cells containing uranium mill tailings. Increasing trends in contamination directly downgradient of the disposal cell may indicate seepage of site contaminants from the cell into the groundwater. On the other hand, the trend may be the result of changes in background quality upgradient from the disposal site, so it will be important to monitor upgradient wells for trends in water quality as well.

Upon removal or stabilization of the tailings at former processing sites, a plan for groundwater restoration will be implemented. The type and intensity of restoration required at a site will depend, in part, on the observed rate of decline in groundwater contamination since tailings removal or stabilization. Risk assessors examine groundwater quality data for trends over time to forecast future, as well as current, health risks associated with a contaminated aquifer. To verify that the aquifer restoration activities are accomplishing their intended purpose, trends in groundwater quality over time must be quantified and compared to trends projected by theoretical groundwater models.

Trend analysis of UMTRA Project groundwater quality data is not being done routinely because many sites have only 3 to 5 years of data. At most sites, surface activities are being finalized with groundwater restoration activities in the planning stages. But as surface activities wind down and sites enter the groundwater restoration and long-term monitoring phases, the information provided by trends analysis will be critical to proper site management.

This report begins by presenting the statistical concepts that are necessary for understanding trend analysis (Section 2.0). The types of patterns most likely to occur in UMTRA data sets are emphasized. Two general approaches to analyzing data for trends are proposed in Section 3.0 and recommendations are given to assist UMTRA Project staff in selecting an appropriate method for their site data. Trend analysis is much more difficult when data contain values less than the reported laboratory detection limit. The complications that arise are explained in Section 4.0. This report also discusses the impact of data collection procedures on statistical trend methods and offers recommendations to improve the efficiency of the methods and reduce sampling costs (Section 5.0). Section 6.0 contains guidance for determining how many sampling rounds might be needed by statistical methods to detect trends of various magnitudes. This information could be useful in planning site monitoring activities.

This report limits its discussion to univariate trends analysis (i.e., evaluation of changes over time in the level of one constituent in one well). Intelligent use of univariate trends analysis will greatly benefit the UMTRA Project. However, in the future, the Project may also need to consider methods that combine information on many constituents and many

wells to characterize trends over time for the aquifer as a whole. Multivariate ("many variable") trends analysis is an extension of the univariate analysis presented in this report, and relies on many of the same statistical principles. However, the subject is conceptually and mathematically more difficult to understand, computationally more difficult to perform, and may require additional and better quality data than the Project currently has for most sites.

2.0 STATISTICAL CONCEPTS FOR TREND ANALYSIS

At any fixed point in time, the concentration of a chemical in a well varies around some average value. The word "trend" usually refers to a change over time in this average value.

2.1 KINDS OF TRENDS

Many different kinds of trends in average value are possible (Figure 2.1). A "step trend" occurs if data collected before a specific time have a distinctively different average than data collected after that time. A step trend might result from a sudden release of a contaminant into the groundwater. A "monotonic trend" occurs if the average value moves consistently in one direction (strictly up or strictly down) over time. For example, a contaminated well is expected to show a long-term monotonic decline in contaminant concentration levels once the source of contamination is eliminated by surface remediation.

Long-term surveillance of UMTRA Project disposal sites includes statistical methods designed to warn of a sudden release of contaminants from the cell into the groundwater. These methods should detect a step trend; however, they are insensitive to gradual changes over time in the level of a contaminant. Therefore, this report focuses on monotonic trends.

Monotonic trends may be linear or curved. A linear trend shows a constant rate of change in concentration level over time. An exponential trend is an example of a curved monotonic trend. In exponential trends, the rate of change in concentration level at any time is proportional to the level at that time.

2.2 ERROR DISTRIBUTIONS

A measurement of water quality almost always differs from the true well average at the time the water sample was collected. Besides natural variation within the well itself, errors are introduced in the sampling process, in handling and transport of the water sample, and in the laboratory analytical procedures that produce the measurement. Errors may also occur in recording the measurement or in transferring the measurement into a computer data base.

If the entire process, from sampling to computer data entry, could be repeated over and over again, simultaneously in time, the net errors (recorded measurement minus true well average) would form a distribution of values. This is called the "error distribution" for the sampling period. Any measurement obtained from a water sample represents the true well average at the time the sample was taken, plus one error selected at random from the error distribution of that sampling period. The error distributions are responsible for the noise or scatter of data common to time-concentration plots.

No
Trend

Step
Trend

Linear
Monotonic
Trend

Curved
Monotonic
Trend

FIGURE 2.1
TYPES OF TRENDS IN CONCENTRATION OVER TIME

Three aspects of error distributions—their mean value, the amount of variability they contain, and their shape—impact trend analysis.

2.2.1 The mean value of an error distribution and bias

If the mean value of an error distribution is positive, then measurements are, on the average, larger than the true well average. Equivalently, a negative mean value signifies that measurements are, on average, too small. In either case, the errors produce a "bias" in the measurements. The unbiased case, corresponding to a mean value of zero, requires that positive and negative errors of the same magnitude occur equally often.

2.2.2 Variability of error distributions

The variability of an error distribution describes how spread out the errors may be from one another. The "standard deviation" is a measure that quantifies the amount of variability of a distribution. The standard deviation equals zero if errors are always exactly the same size; otherwise, the standard deviation is a positive quantity. If errors vary only slightly from one another, the standard deviation of the error distribution will be small. If errors vary a great deal from one another, the standard deviation will be large.

2.2.3 Shape of error distributions

Two error distributions with the same mean value and standard deviation may still differ in distributional shape. Shape relates to the entire spectrum of values an error may assume and the relative frequencies with which those values occur. An error distribution is "unimodal" if the value of the error is relatively more likely to occur within one particular interval than within any other interval of the same width. The error distribution is "symmetrical" around this one mode if the relative frequencies of values are equal on either side of the mode. An example of a symmetrical and unimodal distribution is the normal curve, characteristically bell-shaped. Not all bells represent normal curves, however. If the extremes of a bell extend in both directions further than the normal curve allows, the distribution is said to exhibit "positive kurtosis." If the distribution is unimodal and one extreme of the distribution extends further than the other extreme does, the distribution is "skewed."

2.2.4 Changes in error distributions over time

Sudden modifications or gradual changes in data collection procedures over time may alter the error distributions of subsequent sampling rounds. Examples include improvements in a laboratory analytical procedure, modifications in water sampling methods, or changes in personnel. Externally induced changes over time in the error distributions are superimposed on changes over time in average well

concentration. It is impossible to distinguish between the two by examining the measurement data of a single well. For example, a trend in the amount and direction of bias over time will appear on a time-concentration plot as a trend in average well concentration.

In addition to being affected by external forces, the magnitude of error in a measurement is often dependent in some way on the existing concentration. If so, changes over time in average well concentration would be accompanied by changes in the bias or the variability of the measurements.

Figure 2.2 illustrates a number of error distributions to clarify the concepts of mean value, variability, and shape. Figure 2.3 presents some realistic examples of changes that might occur over time in error distributions when water sampling continues for several years. Figure 2.4 illustrates the result of imposing random errors from various error distributions onto a linear trend.

2.3 HYPOTHESIS TESTS

Trend analysis can take the form of detection or estimation. For trend detection, data are analyzed to decide whether or not a trend in average well concentration exists. Trend detection produces a yes or no answer. Proper site management may further require a number that quantifies the rate of change in average well concentration over time. This represents trend estimation and will be discussed in Section 2.4.

2.3.1 The null and alternative hypotheses in trend analysis

Hypothesis testing uses measured concentration data over time to decide which of two hypotheses about the underlying trend is true. The two hypotheses are called H_0 and H_a .

The H_0 is the null hypothesis and usually contains a statement expressing the "status quo" condition of the well. Depending on the application, H_0 represents "no interesting trend" and "no change in management indicated." In hypothesis tests, inconclusive evidence favors H_0 . The H_a , the alternative hypothesis, represents a statement about the trend that is of interest to the data analyst. A decision that H_a is true usually leads to a specific action. Strong evidence from data is needed to decide in favor of H_a .

For UMTRA Project sites, the following three different arrangements of H_0 and H_a are possible.

1. H_0 : no positive trend exists (trend is zero or negative).
 H_a : there is a positive trend.

Example: Remedial action will be taken if time-concentration data provide clear evidence that contamination is increasing in an aquifer.

FIGURE 2.2
ERROR DISTRIBUTION

FIGURE 2.3
CHANGES IN ERROR DISTRIBUTION OVER TIME

Normal error distribution
Low variation

Normal error distribution
High variation

Normal error distribution
Variation increasing with level

Kurtosis

Positive bias in errors
Bias decreasing over time

FIGURE 2.4
ERROR DISTRIBUTION IMPOSED ON A REAL TREND IN AVERAGE CONCENTRATION

2. H_0 : no negative trend exists (trend is zero or positive).
 H_a : there is a negative trend.

Example: Conclusive evidence is needed that passive flushing is adequate to produce decreases in contamination of an aquifer.

3. H_0 : no trend.
 H_a : a trend exists (either positive or negative).

Example: Background wells will be monitored for changes, either positive or negative, in the groundwater upgradient from the UMTRA Project site.

2.3.2 H_0 and H_a , level of significance, and Type 1 errors

To decide between the hypotheses, the time-concentration data are summarized into a single number called the "test statistic." Typically, large values of the test statistic lend support for H_a and near zero values of the test statistic favor H_0 .

value of
test statistic

If the value of the test statistic exceeds a critical point, represented by c on the sketch above, then H_a is the appropriate decision, and the result is considered statistically significant. To the left of c is a region where support for H_a is inadequate or inconclusive, and H_0 is favored.

The critical point for a hypothesis depends on the level of significance selected for the hypothesis test. The level of significance is the probability that random variation, acting alone, could cause a test statistic to exceed " c ". Whenever that happens, the data analyst will make the wrong decision by concluding H_a is true when, in fact, H_0 is true. This wrong decision is called a "Type 1" error or a false positive result. The level of significance of a test is the probability of making a Type 1 error, or the probability that chance patterns in the error distributions could ever, by themselves, cause H_a to be wrongly chosen over H_0 .

Type 1 errors are usually costly. For example, remedial action could unnecessarily be initiated on an aquifer if a Type 1 error is made. Therefore, hypothesis tests specify a low value for the level of significance such as 0.05 or even 0.01. The value of the critical point " c " reflects the choice of the level of significance. Generally, if it is important to avoid making a Type 1 error, " c " is increased so that more evidence is required for statistical significance.

2.3.3 Type 2 error and power

If there really is a trend in average well concentration, the hypothesis test should decide H_a is true; however, it often does not. If the test statistic falls below the critical point, the decision must be made that H_0 is true, or at least that the results are inconclusive. This kind of decision error is called "Type 2," or a false negative result.

The probability of making a Type 2 error depends on the strength of the underlying trend relative to the variability around the trend. A very slow change in concentration will probably go undetected for years. Even rapid changes in concentration level may not be detectable if random variation in the measurements is excessive.

Type 2 errors are more frequent if low levels of significance are used because protection against Type 1 errors is achieved by increasing the critical point. For example, a real trend is more likely to go undetected if the hypothesis test is done at a 0.01 level of significance than if the larger value 0.05 is used.

2.4 TREND ESTIMATION

A number that quantifies the rate of change over time in average well concentration is called the "trend slope." A trend slope assumes a linear trend in which the rate of change over time in average concentration is constant and equal to the slope of the time-concentration line. For curved trends, the slope changes constantly and a single number is inadequate to quantify the trend.

Throughout the report, the true slope of the trend in average well concentration is symbolized as β . More than one statistical method is available to estimate β from a set of data. The general principles of sampling distributions and confidence intervals discussed in the next subsection apply to all slope estimators.

2.4.1 Sampling distribution of a trend estimator

The error distributions of the concentration measurements introduce error into the estimation process. In fact, if many replicate measurements for each time period could be obtained, a slope estimator would produce an entire distribution of slope estimates. This theoretical distribution is called the "sampling distribution" of the slope estimator.

The mean value of the sampling distribution relates to bias. A trend estimator that tends to overestimate the true slope or, alternatively, to underestimate the true slope, is said to be a "biased estimator." The mean value of the sampling distribution of a biased estimator differs from β . An "unbiased estimator" is one without systematic tendencies; its mean value equals β .

The variability of a sampling distribution relates to an estimator's efficiency. The standard deviation of the sampling distribution is called the "standard error" of the estimator. A large standard error means that the estimator frequently produces very inaccurate estimates, and so is not very reliable. Optimally, an estimator is unbiased and has a small standard error. Such an estimator usually produces very accurate estimates of the slope.

2.4.2 Confidence intervals

Confidence intervals are used to quantify the possible error in an estimate. A confidence interval consists of a range, or interval, of values, together with a stated level of confidence that the true slope β lies within the interval.

For example, suppose β represents the annual change in average concentration of nitrate in a well. Based on a set of data, an estimator produces the point estimate $b = 50$ mg/L. The 95-percent confidence interval for β is reported as 50 ± 20 mg/L or, (30 mg/L, 70 mg/L). The interpretation of this interval is, firstly, that nitrate concentration is definitely increasing over time in this well. Secondly, the rate of change in average concentration could be as low as 30 mg/L or as high as 70 mg/L per year.

The concept of level of confidence is related to the sampling distribution of an estimator. Of all the estimates an estimator could possibly produce, the vast majority are close to the true slope trend β . Only 5 percent are so far away that a 95-percent confidence interval constructed around them will fail to capture β somewhere within the interval. Occasionally, a 95-percent confidence interval will not contain the true slope; however, that will happen only about 5 percent of the time. A data analyst anticipates that approximately one out of every twenty 95-percent confidence intervals will be misleading.

The level of confidence is chosen by the data analyst. It should be close to 100 percent to produce a high degree of trust in the statistical results. However, increasing levels of confidence are always accompanied by wider confidence intervals. Wide intervals offer an extensive range of possible values for the true slope; consequently, the true slope has not been very precisely estimated. For the nitrate example above, a 99 percent level of confidence might produce an interval of (25mg/L, 75mg/L). The effect of increasing the level of confidence from 95 percent to 99 percent is greater confidence in a less precise result. A general rule of thumb is to use a level of confidence no less than 90 percent and no more than 99 percent. If the resulting confidence intervals are too wide to be informative, more data are needed.

3.0 STATISTICAL TREND METHODS

The most promising statistical methods for trend analysis of UMTRA Project water-quality data are parametric linear regression and nonparametric methods. These methods are described briefly below; more complete details of the methods are provided by Helsel and Hirsch in their Statistical Methods on Water Resources (1991).

The data are assumed to be measurements of the level of a constituent at "n" different times in the same well. Symbolically, the data can be described as follows:

$$(T_1, Y_1), (T_2, Y_2), \dots, (T_n, Y_n)$$

where T represents time and Y represents the measured level of the constituent. The data are ordered in time, so that $T_1 < T_2 < \dots < T_n$. The time measurements can be recorded as days, months, or years. It is convenient to subtract the first time T_1 from all measurements, so that $T_1 = 0$ and T_k represents the elapsed time from the first to the kth measurement.

3.1 LINEAR REGRESSION

Linear regression is based on a user-specified model of the trend in average well concentration over time. The model includes one or more unknown constants or parameters. The parameters are estimated from the data to maximize the agreement between the data and the model.

3.1.1 The simple linear regression model

A simple yet powerful linear regression model follows:

$$Y = \alpha + \beta T + \text{error}.$$

This model assumes the trend in average well concentration is linear over time, so that the true average well concentration at time T equals $\alpha + \beta T$. The measurements are assumed to contain errors. According to this model, the error distributions do not change over time and do not add a bias into the measurements.

The parameters for this model are α , β , and σ . The first parameter, α is called the y-intercept and represents the average well concentration at time 0. The slope of the trend is β . The third parameter, σ , represents the standard deviation of the error distribution. Trend analysis usually focuses on the slope β ; however, σ is also important as an indicator of how much error may be present in measurements. The y-intercept α is not generally of interest.

The parameters β and σ are unknown quantities that are estimated from the data via a process called "ordinary least squares." The estimates are denoted b and s. Different letters are used to distinguish the unknown parameters from their estimates.

To construct a confidence interval or to conduct a hypothesis test about the trend slope, β , linear regression assumes that the error distributions are normal. The computations required to produce estimates, confidence intervals, or hypothesis tests are given in most introductory statistics textbooks and will not be described here.

3.1.2 Other linear regression models

It is not necessary to assume that the trend is a straight line over time. Some curved models can be used with least squares methods. However, least squares methods are primarily limited to models for which the parameters appear in different terms of the model and act as multipliers of time or of a function of time. The word "linear" in the expression "linear regression models" reflects the linear placement of the parameters, not the type of trend itself. An example of a linear model that allows for a curved trend over time is as follows:

$$Y = a + \beta T^2 + \text{error}$$

This model states that the trend in average concentration is quadratic over time and may be appropriate if the time-concentration plot of the data is parabolic. To estimate the parameters in the model, all computations would involve the square of time rather than time itself. Except for this change, procedures are identical to those for the simple linear regression model.

An extension of parametric least squares allows the standard deviation of the error distributions to change over time. Measurement errors are commonly proportional to the average well concentration being measured. When well concentrations of a constituent are high, the absolute errors in the measurements may be quite large; when the concentration is low, the absolute errors are typically small. Percentage-wise, however, errors often tend to be larger at low and high concentrations and least in the middle, where analytical instruments and methods are optimized. Weighted least squares is a procedure that can adjust for unequal variation in the measurements. The data are assigned weights so that data thought to be less reliable have reduced influence on the slope estimate.

3.1.3 Checking model assumptions

Statistical software packages for linear regression include graphic tools for checking model assumptions. "Residuals" from a model are the differences between the measurements and the values predicted by the model. For the simple linear regression model, the residual at time T_k is $Y_k - (a + bT_k)$. If the model assumptions are all correct, the residuals reflect only the random measurement errors and an x-y plot of time versus residuals will show a random scatter of points. However, if a linear trend model is incorrectly chosen as a model for data that exhibit a curve, the residual plot will show a trend, flagging that an inadequate model was used on the data. The residual plot will also indicate if the variability of errors is changing over time.

Another useful graphic tool is a "boxplot" of the residuals. A boxplot is a picture of the distributional shape of the residuals and provides information on the shape of the actual error distribution. A boxplot can be used, for example, to see if the assumption of a normal error distribution is reasonable. Examples of boxplots are provided in Figure 3.1.

3.2 NONPARAMETRIC METHODS

Nonparametric methods of trend analysis rely on a much weaker set of assumptions about the type of trend than are required in linear regression. Nonparametric methods distinguish between only three possible trends in average well concentration. An increasing trend is one for which measurements later in time are likely to be larger than measurements earlier in time. A decreasing trend is similar, except that later measurements are likely to be smaller than earlier measurements. If neither case applies, no trend exists.

Few assumptions are required for the error distributions affecting the measurements. Nonparametric methods implicitly assume that any bias in the measurements is constant over time. The standard deviations of the error distributions are allowed to vary monotonically over time.

Nonparametric methods include the Mann-Kendall hypothesis test (Mann, 1945) and the Kendall-Theil trend estimator (Theil, 1950). Both procedures look at all possible pairs of measurements (T_i, Y_i) and (T_j, Y_j) where $j > i$, so that Y_j represents a measurement taken later in time than Y_i . If the data set contains "n" concentration measurements, then $n(n-1)/2$ measurement pairs exist.

The Mann-Kendall hypothesis test assigns to each pair a score of 1, -1, or 0 according to whether the later measurement is greater than, less than, or equal to the earlier measurement. The sum of all these scores is the test statistic used to decide if a statistically significant trend exists. Large positive sums are evidence of a positive trend, whereas, a large negative sum suggests the trend is negative. A sum near zero is likely to occur if no trend exists. Notice that the actual measurements are not used in computing the Mann-Kendall test statistic. The test statistic depends only on the rankings of the "n" measurements (smallest measurement has rank 1, ... largest measurement has rank n).

The Kendall-Theil slope estimator calculates the slope of the line connecting the two points (T_i, Y_i) and (T_j, Y_j) , where:

$$\text{SLOPE} = (Y_j - Y_i)/(T_j - T_i).$$

Each pair of measurements produces a slope. The median of all the slopes becomes the estimate of the slope of the trend. The median is the middle value of an ordered set of numbers.

If the number of data values is small, Table B8 in Helsel and Hirsch (1991) can be used to test for the existence of a trend and/or to construct a confidence interval

Prototype
Normal
Boxplot

Well Number

901

904

910

0

914

920

921

0

FIGURE 3.1
BOXPLOTS OF NITRATE MEASUREMENTS AT SIX BACKGROUND WELLS
TUBA CITY, ARIZONA

for the trend slope. Specific instructions on how to use the table are provided in the reference. The table is supplemented by approximations that must be used on data sets consisting of more than 10 measurements.

3.3 EXAMPLE

Least squares regression and nonparametric trend analysis will often provide similar results. But occasionally the methods differ in their conclusions. An example is provided here to illustrate the differences between the methods and their interpretations.

The example is from a monitor well at the Tuba City, Arizona, former processing site. The tailings were stabilized in place in 1988. Well 906 is located directly downgradient from the cell in the contamination plume. Ten uranium measurements were collected on well 906 between 1988 and 1991 and are reproduced in Table 3.1. The sampling dates were transformed into years elapsed since the first measurement was taken in May 1988, and the measured concentrations are in milligrams per liter.

Table 3.1 Uranium concentration in well 906 during 1988-91, Tuba City, Arizona

Sampling date	Years elapsed	U mg/L
5/12/88	0.000	0.958
9/14/88	0.342	0.720
12/13/88	0.590	0.636
6/27/89	1.126	0.990
2/22/90	1.784	0.590
6/29/90	2.132	0.554
11/1/90	2.474	0.661
3/14/91	2.838	0.594
6/21/91	3.110	0.531
9/19/91	3.351	0.502

For the first few years after a tailings pile is covered, it cannot be known for certain whether groundwater contamination will increase or decrease. An increase could result from a decrease in the recharge of uncontaminated water into the aquifer or from an increase in the transient drainage from a pile. A decrease might occur as the plume moves downgradient. Therefore, the hypothesis of interest for this well is whether a trend exists, either increasing or decreasing. The plot of time versus concentration in Figure 3.2 suggests that the trend is, in fact, negative.

Parametric Linear Regression

$$U = 0.855 - 0.102 \text{ Time}$$

Predictor	Coef	Stdev	t-ratio	p
Intercept	0.85451	0.07350	11.63	0.000
Slope	-0.10195	0.03483	-2.93	0.019

$$s = 0.126$$

95% Confidence Interval for Slope is:

$$-0.102 - (2.3060) (.03483) , -0.102 + (2.3060) (.03483)$$

$$(-0.182 , -0.022)$$

99% Confidence Interval for Slope is:

$$-0.102 - (3.3554) (.03483) , -0.102 + (3.3554) (.03483)$$

$$(-0.219 , 0.015)$$

FIGURE 3.2
PARAMETRIC LINEAR REGRESSION ANALYSIS OF URANIUM DATA

3.3.1 Parametric linear regression analysis

These results of parametric linear regression analysis of the uranium data are presented in Figure 3.2. The method estimates that average uranium concentration is decreasing at a rate of 0.102 mg/L per year ($b = -0.102$). The estimated standard deviation of the error distribution is 0.126 mg/L.

The hypothesis test about the slope trend produces a p-value of 0.019. This means that the slope is statistically significant using any level of significance greater than 0.019. The slope is significant at a 0.05 level of significance but not at a 0.01 level of significance.

The parametric 95-percent confidence interval for the true trend slope is $-0.102 \pm 0.080 = (-0.182, -0.022)$. It can be asserted with 95-percent confidence that average well concentration is dropping at an annual rate of somewhere between 0.022 and 0.182 mg/L per year. If even more confidence is desired, the 99-percent confidence interval of $-0.102 \pm 0.117 = (-0.219, 0.015)$ can be used. The 99-percent confidence interval contains positive as well as negative values, reflecting the fact that the hypothesis test at a 0.01 level of significance was unable to establish that a trend exists.

3.3.2 Nonparametric trend analysis

The steps involved in nonparametric trend analysis of the uranium data are shown in Table 3.2. With 10 data values in this example, there are $(10)(9)/2 = 45$ possible pairs of data points (time, concentration). The first pair are (0.000, 0.958) and (0.342, 0.720). For this pair, the later uranium measurement is smaller than the earlier one, so a score of -1 is assigned to the pair. When similar assessments are made on the remaining 44 pairs of data points, 8 of the pairs are such that the later uranium measurement is larger than the earlier one; these pairs receive a Mann-Kendall score of 1. For the remaining pairs, the later measurement is smaller than the earlier one and are assigned a score of -1. The sum of all 45 scores is $8(1) + (37)(-1) = -29$. This is the Mann-Kendall test statistic. The p-value of the result is 0.0092; therefore, the trend is deemed statistically significant at any level of significance larger than 0.0092. In particular, the trend is significant at both 0.05 and 0.01 levels of confidence.

Also listed in Table 3.2 are the slopes of the lines correcting each pair of data points. For the first pair of data points, the slope is calculated as $(0.720 - 0.958)/(0.342 - 0.000) = -0.695$. The median of the 45 slopes is the 23rd value when the slopes are ordered from smallest to largest or vice versa; the median is -0.093. Therefore, the nonparametric slope estimate of this data set is -0.093. According to Table B-8 of Helsel and Hirsch (1991), the lower and upper 95 percent confidence limits for β are the 12th and 34th ordered slopes, or (-0.205, -0.309). The 99 percent confidence interval consists of all real numbers between the 9th and the 37th ordered slopes, or (-0.231, -0.019).

TABLE 3.2
NONPARAMETRIC TREND ANALYSIS OF URANIUM DATA

J	I	(T _j , Y _j)	(T _i , Y _i)	SCORE	SLOPE
2	1	(0.342, 0.720)	(0.000, 0.958)	-1	-0.695
3	1	(0.590, 0.636)	(0.000, 0.958)	-1	-0.547
3	2	(0.590, 0.636)	(0.342, 0.720)	-1	-0.341
4	1	(1.126, 0.990)	(0.000, 0.958)	1	0.028
4	2	(1.126, 0.990)	(0.342, 0.720)	1	0.345
4	3	(1.126, 0.990)	(0.590, 0.636)	1	0.659
5	1	(1.784, 0.590)	(0.000, 0.958)	-1	-0.206
5	2	(1.784, 0.590)	(0.342, 0.720)	-1	-0.090
5	3	(1.784, 0.590)	(0.590, 0.636)	-1	-0.039
5	4	(1.784, 0.590)	(1.126, 0.990)	-1	-0.608
6	1	(2.132, 0.554)	(0.000, 0.958)	-1	-0.190
6	2	(2.132, 0.554)	(0.342, 0.720)	-1	-0.093
6	3	(2.132, 0.554)	(0.590, 0.636)	-1	-0.053
6	4	(2.132, 0.554)	(1.126, 0.990)	-1	-0.434
6	5	(2.132, 0.554)	(1.784, 0.590)	-1	-0.103
7	1	(2.474, 0.661)	(0.000, 0.958)	-1	-0.120
7	2	(2.474, 0.661)	(0.342, 0.720)	-1	-0.028
7	3	(2.474, 0.661)	(0.590, 0.636)	1	0.013
7	4	(2.474, 0.661)	(1.126, 0.990)	-1	-0.244
7	5	(2.474, 0.661)	(1.784, 0.590)	1	0.103
7	6	(2.474, 0.661)	(2.132, 0.554)	1	0.312
8	1	(2.838, 0.594)	(0.000, 0.958)	-1	-0.128
8	2	(2.838, 0.594)	(0.342, 0.720)	-1	-0.050
8	3	(2.838, 0.594)	(0.590, 0.636)	-1	-0.019
8	4	(2.838, 0.594)	(1.126, 0.990)	-1	-0.231
8	5	(2.838, 0.594)	(1.784, 0.590)	1	0.004
8	6	(2.838, 0.594)	(2.132, 0.554)	1	0.057
8	7	(2.838, 0.594)	(2.474, 0.661)	-1	-0.184
9	1	(3.110, 0.531)	(0.000, 0.958)	-1	-0.137
9	2	(3.110, 0.531)	(0.342, 0.720)	-1	-0.068
9	3	(3.110, 0.531)	(0.590, 0.636)	-1	-0.042
9	4	(3.110, 0.531)	(1.126, 0.990)	-1	-0.231
9	5	(3.110, 0.531)	(1.784, 0.590)	-1	-0.044
9	6	(3.110, 0.531)	(2.132, 0.554)	-1	-0.024
9	7	(3.110, 0.531)	(2.474, 0.661)	-1	-0.205
9	8	(3.110, 0.531)	(2.838, 0.594)	-1	-0.232
0	1	(3.351, 0.502)	(0.000, 0.958)	-1	-0.136
0	2	(3.351, 0.509)	(0.342, 0.720)	-1	-0.072
0	3	(3.351, 0.509)	(0.590, 0.636)	-1	-0.049
0	4	(3.351, 0.509)	(1.126, 0.990)	-1	-0.219
0	5	(3.351, 0.509)	(1.784, 0.590)	-1	-0.056
0	6	(3.351, 0.509)	(2.132, 0.554)	-1	-0.043
0	7	(3.351, 0.509)	(2.474, 0.661)	-1	-0.181
0	8	(3.351, 0.509)	(2.838, 0.594)	-1	-0.180
0	9	(3.351, 0.509)	(3.110, 0.531)	-1	-0.120

-29

SLOPES SORTED

-0.695	-0.608	-0.547	-0.434	-0.341	-0.244	-0.232	-0.231
-0.231	-0.219	-0.206	-0.205	-0.190	-0.184	-0.181	-0.180
-0.137	-0.136	-0.128	-0.120	-0.120	-0.103	-0.093	-0.090
-0.072	-0.068	-0.056	-0.053	-0.050	-0.049	-0.044	-0.043
-0.042	-0.039	-0.028	-0.024	-0.019	0.004	0.013	0.028
0.057	0.103	0.312	0.345	0.659			

The results of the two methods of trend analysis are placed side by side in Table 3.3.

Table 3.3 Comparison of parametric and nonparametric analyses of uranium data

	Parametric	Nonparametric
Slope estimate	- 0.102	- 0.093
Hypothesis test results		
0.05 level of significance	significant	significant
0.01 level of significance	not significant	significant
Confidence interval for slope		
95 percent confidence	(-0.182, -0.022)	(-0.205, -0.039)
99 percent confidence	(-0.219, 0.015)	(-0.231, -0.019)

The parametric and nonparametric results are very similar. The two slope estimates differ by only 0.009 mg/L. At the 0.05 level of significance, the methods produce the same conclusion. The 95 percent confidence intervals are also almost exactly the same width, indicating the same efficiency or precision of results.

Unfortunately, if the hypothesis test is conducted at the 0.01 level of confidence, the methods yield different conclusions. The nonparametric test result indicates a significant trend while the parametric result is inconclusive. Recognizing that the p-value for the parametric test is 0.019, which is only slightly larger than 0.01, it seems safe to conclude that uranium concentration is decreasing over time. However, this example illustrates the possible impact of the choice of statistical method on the results of trend analysis.

3.4 COMPARISON OF PARAMETRIC AND NONPARAMETRIC TREND METHODS

3.4.1 Relative efficiency of the two methods

Parametric linear regression relies on a set of assumptions concerning the type of trend that exists in average well concentration and the shape of the error distributions affecting concentration measurements. If all the assumptions made by linear regression are true, then the parametric procedure has greater efficiency than the nonparametric method. The parametric hypothesis test is more likely to detect a trend in average well concentration and the parametric confidence intervals will usually be narrower than the nonparametric counterparts. The difference in the efficiencies of the two methods is approximately 10 percent; that is, the nonparametric method requires approximately 10 percent more data to be comparable in efficiency to the parametric method.

The assumption of normal error distributions is critical to the efficiency of parametric procedures. Statistical research has shown that parametric linear regression loses efficiency when the error distributions are skewed or kurtotic (Sen, 1968). These non-normal error distributions produce extremely large errors more frequently than the normal distribution. The unusual measurements may appear as "outliers" on the time-concentration plot of the data. Outliers are measurements that do not fit the pattern exhibited by the rest of the data. Skewed distributions produce outliers on one side of the general trend, while kurtotic error distributions can produce outliers that lie above or below the general trend. When outliers occur, the least squares estimate of the standard deviation of the error distribution is inflated. Since this standard deviation is used to calculate the parametric test statistic and confidence interval, they are also affected. Inconclusive test results and wide confidence intervals may result.

The nonparametric procedures are affected by outliers to a much lesser extent than the parametric procedure. The nonparametric procedure is more efficient than the parametric procedure when the error distributions are skewed or kurtotic. One recent study showed the nonparametric procedure to be up to 20 percent more efficient than parametric least squares regression when the skewness and kurtosis of the error distributions were so slight as to be visually indistinguishable from normality (Hirsch et al., 1991).

UMTRA Project water-quality data are likely to have skewness or kurtosis. To illustrate this, nitrate measurements from six background wells at Tuba City, Arizona, are displayed with boxplots in Figure 3.1. A prototype of a normal boxplot is also shown in the figure for comparison. Nitrate levels are stable in background, so the variation between measurements at a well is strictly the result of the error distributions.

The "*" on the boxplots are measurements that the normal distribution might produce approximately 3 percent of the time. The "O" indicates measurements that are almost impossible under the assumption of normality. The "Os" are extreme outliers. A single "O" on a boxplot is sufficient warning of non-normal data.

Trend analysis of the nitrate data from well 921 shows the effect of outliers. As Figure 3.1 indicates, well 921 has one extremely large measurement. The parametric 95-percent confidence interval estimate of the slope trend is $(-4.303, 1.897)$ mg/L per year. The width of this interval is 6.200. Nonparametric trend analysis produces a 95 percent confidence interval of $(-4.502, 0.750)$ mg/L per year. The nonparametric interval width is 5.252, which is 15 percent narrower than the parametric interval. It should be noted, however, that both parametric and nonparametric intervals contain positive and negative values. Neither procedure can determine whether the trend over time is positive, negative, or zero.

Unfortunately, there is no limit on the loss of efficiency of parametric methods as more and more extreme outliers are introduced into the measurements. Figure 3.3 presents a hypothetical data set for which nine consecutive years of measurements lie exactly on a line with a slope of -0.100. On the tenth year, an extreme error

Parametric Linear Regression

$$Y = 0.724 + 0.036 \text{ Time}$$

Predictor	Coef	Stdev	t-ratio	p
Intercept	0.7236	0.3194	2.27	0.053
Slope	0.00364	0.05983	0.06	0.953

$$s = 0.5435$$

95% Confidence Interval for the Slope is:

$$0.00364 - (2.3060)(.05983), 0.00364 + (2.3060)(.05983)$$

$$(-0.134, 0.146)$$

Non-Parametric Trend Analysis All Possible Slopes

-0.10000	-0.10000	-0.10000	-0.10000	-0.10000	-0.10000	-0.10000
-0.10000	-0.10000	-0.10000	-0.10000	-0.10000	-0.10000	-0.10000
-0.10000	-0.10000	-0.10000	-0.10000	-0.10000	-0.10000	-0.10000
-0.10000	-0.10000	-0.10000	-0.10000	-0.10000	-0.10000	-0.10000
-0.10000	-0.10000	-0.10000	-0.10000	-0.10000	-0.10000	-0.10000
-0.10000	0.11111	0.13750	0.17143	0.21667	0.28000	0.37500
0.53333	0.85000	1.80000				

FIGURE 3.3
EXAMPLE OF LACK OF EFFICIENCY OF PARAMETRIC PROCEDURE WHEN
AN EXTREME OUTLIER IS PRESENT

affects the measurement. The parametric method estimates the slope to be positive due to the influence of that last measurement. The 95 percent confidence interval for the slope trend is (-0.134, 0.142). The most that can be said about this procedure is that it succeeds in capturing the true slope (-0.10) in the confidence interval. The nonparametric method estimates the slope to be -0.10 and the confidence interval is (-0.10, -0.10); in other words, the confidence interval has width zero. Essentially, the error in the last measurement had no influence at all in the nonparametric procedure.

In conclusion, the parametric linear regression methods are slightly more efficient than nonparametric methods under ideal (normal) conditions; however, they may be much less efficient under the real life conditions that are likely to exist in UMTRA Project data sets. Parametric procedures have supplementary tools that will detect kurtosis and down-weight the influence of outliers, but nonparametric methods have this advantage built into the method. With the nonparametric procedure, a reliable estimate can be expected with very little user input and interface.

3.4.2 Other considerations in selecting a trend method

Parametric linear regression methods have advantages when more complicated models are needed to describe changes in concentration over time. For example, if the trend over time is obviously curved, an alternate mathematical model can be employed by linear regression to describe the curve. Nonparametric methods do not have the ability to distinguish between a line and a curve. In fact, if concentration increases during one time interval then changes direction and decreases for a subsequent time interval, nonparametric analysis may view the situation as indicating no trend at all.

Parametric linear regression can also utilize other variables that are known or suspected to influence the measured concentration of a chemical. Examples include pH and water level. High water levels, for example, may be associated with a temporary dilution of the groundwater contaminant. If water level is correlated with contamination measurements, changing water levels over time contribute to the variability of the measured concentrations. Linear regression methods can estimate the effect of water level on concentration and subtract its influence from the measurements. The resulting reduction in variability of the time-concentration plot will allow time trends to be more easily seen. Nonparametric methods lack this ability to adjust for covariables.

Practical considerations in selecting a trend analysis method include the level of statistical training required and the availability of computer software to conduct the analyses. Most of the UMTRA Project technical staff received some college training in linear regression methods, but are unfamiliar with nonparametric trend methods. Similarly, linear regression software are widely available, whereas nonparametric software are currently unavailable to Project staff. Consequently, parametric linear regression will continue to be the default method employed on the Project unless training and computer support are provided to encourage use of nonparametric trend analysis methods.

Lastly, the nonparametric test statistic and slope estimator have "discrete" probability distributions. This means that the test statistic and slope estimate can assume only a limited number of values. The probabilities with which these values occur may not allow the analyst to use the level of significance or level of confidence of choice. If, for example, nonparametric analysis is used on a data set consisting of six measurements, the only available levels of confidence greater than 90 percent are 94.4, 98.3, and 99.7. As another example, throughout this report, the nonparametric "95 percent" confidence intervals for data sets containing 10 values were really 95.4 percent confidence intervals, and the "99 percent" intervals were actually 99.1 percent intervals. The discreteness is quite pronounced for small sample sizes, but disappears rapidly as the number of samples increases. Thus, the problem is more of a nuisance than a real consideration in selecting a trend analysis method.

3.5 RECOMMENDATIONS FOR SELECTING A TREND ANALYSIS METHOD

Time versus concentration plots should always be examined prior to statistical analysis of the data. The plots may uncover outliers that do not follow the general trend. An unusual measurement that can be traced to an analytical or recording error should definitely be corrected. Otherwise, the outliers should be retained in the data set.

If the time-concentration plot shows an obvious curve, or if variables in addition to time are potential predictors of concentration, then some combination of parametric and nonparametric procedures may be required. Helsel and Hirsch (1991) provide some examples of this hybrid approach to trend analysis.

Data sets that appear to show a linear trend and have no outliers can be analyzed using either parametric or nonparametric methods. The results should be roughly the same using either method, and the choice may ultimately be determined by the availability of computer software. If outliers are present in the data set, the nonparametric method is recommended. Judicious use of linear regression, combined with residual analysis and weighted least squares, might produce a trend assessment that is comparable or occasionally superior to the nonparametric results; however, the additional time and expertise required are probably unwarranted in most cases.

4.0 TREND DETECTION WITH CENSORED CONCENTRATION MEASUREMENTS

Many UMTRA Project data sets contain measurements that are recorded as being less than a specified laboratory detection limit. These measurements are said to be "censored" at the reported detection limit. The presence of censored values complicates trend analysis because the actual measured concentrations are not available for calculations.

4.1 SINGLE DETECTION LIMITS

If all the undetectable measurements are subject to the same detection limit, one approach is to substitute a number in place of the censored values for trend analysis calculations. The number should be somewhere between zero and the detection limit itself. One-half the detection limit is a common choice for substitution.

Parametric least squares estimates of beta (β) and sigma (σ) may be heavily dependent on which number is substituted for the censored values (Helsel, 1991). Reasons parallel the influence of outliers on parametric regression. A parametric alternative to substitution is a procedure known as TOBIT regression (Hald, 1949). TOBIT regression uses a procedure called "maximum likelihood" to estimate model parameters. The procedure incorporates the available information from censored values without the need for arbitrary substitution of numbers in their place. Unfortunately, TOBIT regression relies on the assumption of normal errors and possesses inefficiencies comparable to linear regression when the error distributions are non-normal. In addition, TOBIT estimators are slightly biased. The bias may be reduced in some cases by using an adjustment developed by Cohn (1988). Overall, TOBIT regression has limited potential use on UMTRA Project data sets for which only one detection limit applies.

Nonparametric methods require substitution of a number in place of the censored values; however, their results are often unaffected by the choice of number. The Mann-Kendall trend test will produce the same result regardless of which number is substituted. The Kendall-Theil slope estimator will be affected in some instances and not in others. A practical solution may be to substitute first zero and then the detection limit in place of the censored data and compare the slope estimates and confidence intervals produced under these two extreme substitution regimes. If the results are comparable, there is little reason for concern. If the results differ materially depending on which value is substituted for the censored values, it is suggested that the analysis be restricted to trend detection only.

4.2 MULTIPLE DETECTION LIMITS

The reported detection limits of chemicals in groundwater can change over time. It is common in UMTRA Project data sets for there to be two to five different detection limits within a 5-year period. Multiple detection limits create enormous problems for trend analysis. Improvements in laboratory analytical procedures

typically result in a reduction in the reported detection limits over time. A decreasing trend in the reported detection limits can easily be mistaken for a trend in concentration if substitution methods are used for the censored values. To illustrate this, suppose that 4 yearly measurements of concentration are recorded as follows:

time	1	2	3	4
detection limit	0.100	0.010	0.010	0.001
measurement	<0.100	0.020	<0.010	0.004

The reported detection limits are decreasing over time. Substitution of these detection limits for the censored values creates the appearance that the concentrations themselves are decreasing over time. Substituting one-half the detection limit does not improve the situation:

time	1	2	3	4
detection limit	0.100	0.010	0.010	0.001
1/2 DL substitution	0.050	0.020	0.005	0.004

The undetectable measurements at time 1 and 3 might, in reality be as shown below:

time	1	2	3	4
detection limit	0.100	0.010	0.010	0.001
measurement	0.001	0.020	0.003	0.004

If this were the case, the trend would appear to be positive with a decimal error in the measurement at time 2.

Two solutions have been proposed in water-quality literature for handling data with multiple detection limits. One approach is to apply the highest detection limit to all the data in the data set. For the example above, the highest detection limit is 0.100, so all four measurements would be considered to be censored at 0.100. Information is lost by this expediency and the procedures that use this approach could suffer a great reduction in efficiency, depending on how many data points are affected. The second approach is TOBIT regression, mentioned earlier in connection with single detection limits. Again, TOBIT regression incorporates the information contained in the censored values as well as the uncensored ones, so no information is discarded or added. For multiple censored data, TOBIT analysis, even applied to data that have non-normal error distributions, might be a more efficient procedure than the censoring scheme described above.

5.0 EFFECT OF DATA COLLECTION METHODS ON TREND ANALYSIS

The efficiency of a statistical trend analysis method is affected by the frequency, timing, and method of data collection and analysis. This section offers guidance on ways to improve the efficiency of trend analysis or to maintain current levels of efficiency at a reduced cost to the UMTRA Project.

5.1 METHODS OF DATA COLLECTION AND ANALYSIS

Water sampling, handling, transport, laboratory analysis, and recording contribute to errors in the measurement of concentrations available for trend analysis. Procedures should be continuously scrutinized for ways to reduce variability because a reduction in the variability of the measurements will increase the efficiency of any statistical trend method, parametric or nonparametric. Increased efficiency translates directly into cost reduction. If the standard deviation of the errors is reduced by 25 percent, then the statistical method will have comparable power to detect a trend in average well concentration with 25 percent fewer sampling rounds. Of course, some errors, such as the natural variation in concentration within the well itself, cannot be controlled or reduced.

Particular attention should be focused on procedures that add a bias into the measured concentration. A measurement bias that is constant over time will not affect trend slope estimators; however, changes in the bias of the measurements over time will add to or subtract from any trend simultaneously occurring in average well concentration. Nevertheless, reductions in bias are desirable. If a procedural change may alter the bias of measurements, data from many wells that would be affected by the change, especially stable background wells, should be examined for similar changes in measured concentration. This information can be used in the trend analysis to adjust for the change in the measurement bias.

5.2 TIMING OF SAMPLING

Many UMTRA Project wells are presently sampled twice a year with a sampling interval of approximately 6 months. If well concentration naturally fluctuates with season, the timing of the sampling rounds can impact the variability of measurements, and, consequently, the efficiency of trend analysis. If, for example, the semiannual sampling rounds are conducted when concentrations are at the highest and lowest points of the yearly cycle, the seasonal variability will contribute to the overall scatter of measurements. From the perspective of reducing variability, it is better for the sampling to occur during the two periods of the year that are located halfway between the annual maximum and minimum. Figure 5.1 shows the advantage of shifting the sampling period by 3 months to miss the extremes of the annual cycle in concentration levels. Alternatively, trend analysis can adjust for seasonal effects provided a consistent sampling interval is maintained every year.

Linear Trend

+

Annual Cycle

=

Well Concentration

High and Low Sampled

+

Random Error

=

Data for Trend Analysis

Midpoint Sampled

+

Random Error

=

Data for Trend Analysis

FIGURE 5.1
EFFECT OF TIMING OF SEMIANNUAL SAMPLING ON TREND ANALYSIS WHEN
CONCENTRATION HAS SEASONAL VARIATION

Groundwater quality data are not likely to exhibit the large seasonal swings typical of surface water; however, some variation during the year could occur in aquifers recharged primarily by snowmelt. At UMTRA Project sites where seasonal variation is suspected, it may be cost effective to temporarily sample wells more frequently to assess the yearly cycle and then design the sampling program accordingly.

5.3 FREQUENCY OF SAMPLING

The efficiency of a trend analysis method increases with the number of measurements in the data set. In addition, the spacing of the measurements impacts efficiency. Increasing the sampling interval will enhance trend analysis by reducing the standard error of the slope estimator. For example, a trend will be more precisely estimated from 10 annual measurements than from 10 semiannual measurements.

The three time-concentration plots in Figure 5.2 demonstrate this phenomenon. The plots show the scatter around a central linear trend as it might appear after 5 years of semiannual sampling, 10 years of annual sampling, and 20 years of biannual sampling. The same linear trend, $Y = 5 + T$, was used for all three plots. Also, exactly the same random errors were superimposed on the central trend; the only difference is the spacing of the data. The width of both parametric and nonparametric confidence intervals for beta (β) are reduced by a factor of 2 for each doubling of the sampling interval.

Increasing the sampling interval will improve trend analysis at no additional cost. Granted, more years of sampling will be needed to detect a trend if sampling is done less frequently; however, doubling the sampling interval will not double the time required for trend detection. The same information can be obtained at a reduced cost by less frequent sampling. Table 6.1 in Section 6.0 quantifies the increased efficiency of annual sampling compared to the current semiannual program.

A. Semi-Annual Sampling for 5 Years (10 Measurements)

Time	True Trend	Random Error	MEAS (Y)
0.5	5.5	-0.02	5.48
1.0	6.0	1.67	7.67
1.5	6.5	-0.77	5.73
2.0	7.0	-0.15	6.85
2.5	7.5	-1.75	5.75
3.0	8.0	-3.58	4.42
3.5	8.5	2.38	10.88
4.0	9.0	-1.86	7.14
4.5	9.5	0.59	10.09
5.0	10.0	-1.11	8.89

B. Annual Sampling for 10 Years (10 Measurements)

Time	True Trend	Random Error	MEAS (Y)
1.0	6.0	-0.02	5.98
2.0	7.0	1.67	8.67
3.0	8.0	-0.77	7.23
4.0	9.0	-0.15	8.85
5.0	10.0	-1.75	8.25
6.0	11.0	-3.58	7.42
7.0	12.0	2.38	14.38
8.0	13.0	-1.86	11.14
9.0	14.0	0.59	14.59
10.0	15.0	-1.11	13.89

FIGURE 5.2
EFFECT OF SAMPLING INTERVAL ON TREND ESTIMATION

C. Bi-Annual Sampling for 20 Years (10 Measurements)

Time	True Trend	Random Error	MEAS (Y)
2.0	7.0	-0.02	6.98
4.0	9.0	1.67	10.67
6.0	11.0	-0.77	10.23
8.0	13.0	-0.15	12.85
10.0	15.0	-1.75	13.25
12.0	17.0	-3.58	13.42
14.0	19.0	2.38	21.38
16.0	21.0	-1.86	19.14
18.0	23.0	0.59	23.59
20.0	25.0	-1.11	23.89

FIGURE 5.2 (CONTINUED)
EFFECT OF SAMPLING INTERVAL ON TREND ESTIMATION

6.0 SAMPLE SIZES NEEDED FOR TREND DETECTION

The probability that a statistical trend method will detect a statistically significant trend in a data set is called its "power." Power depends on many factors, including the level of significance used, the magnitude of the slope trend, the standard deviation of the error distributions, the spacing between sampling rounds, and the total number of measurements in a data set. These issues have been individually discussed throughout the report; in this section, their joint impacts are quantified.

6.1 THE SIGNAL-TO-NOISE RATIO

The power of a method depends on the true slope trend β and the standard deviation of the error distributions σ only by way of their ratio β/σ , which is called the "signal-to-noise" ratio. The signal-to-noise ratio will be large if well concentration is rapidly changing, or if the variability around the trend is small. Statistical procedures have no difficulty detecting such trends from a large data set. The probability of detecting that same strong trend from a small data set is less, primarily because a single chance error can dominate a small data set. The contrast between large sample and small sample statistical power leads naturally to the question of sample size. How many measurements are required in a data set to be ensured that statistical trend analysis will identify a trend if one really exists?

6.2 POWER CALCULATIONS

Table 6.1 shows the net impact of the level of significance, the "signal-to-noise" ratio, the sampling interval and the total sample size on power. The simple linear model described in Section 3.0 is used for the calculations. Because the assumptions required by the model may not hold for some UMTRA Project data sets, the probabilities in the table are slightly optimistic. They should be considered as upper bounds for the statistical power achievable by trend analysis methods.

If only 5 years of semiannual data are available, no statistical method will detect a trend unless that trend is quite strong. Using the conservative 0.01 level of significance to protect against false positives, the test has less than a 1 in 20 (0.048) chance of detecting a trend if the signal-to-noise ratio is 0.25; the power increases only to 0.219 if that ratio doubles to 0.50. To have reasonable power of detecting a trend, the signal-to-noise ratio must be around 1.0. The uranium concentration trend of Appendix A has a signal-to-noise ratio of approximately 1.0.

A modest increase in the level of significance from 0.01 to 0.05 will increase the power of trend analysis to detect a real trend. The power associated with a signal-to-noise ratio of 0.25 increases from 0.048 to 0.165. Prospects improve from 0.219 probability to 0.509 probability of detecting a trend when the signal-to-noise ratio is 0.50.

Table 6.1 Power or probability^a of detecting a statistically significant trend in a water quality variable

No. of years of data	No. of samples per year	True signal-to-noise ratio ^b			
		0.10	0.25	0.50	1.00
5	1	0.056 ^c (0.011)	0.088 (0.019)	0.200 (0.049)	0.574 (0.190)
	2	0.068 (0.015)	0.165 (0.048)	0.509 (0.219)	0.983 (0.844)
10	1	0.123 (0.033)	0.509 (0.219)	0.983 (0.844)	d
	2	0.219 (0.077)	0.876 (0.633)	d	d
15	1	0.328 (0.128)	0.980 (0.863)	d	d
	2	0.626 (0.350)	d	d	d
20	1	0.687 (0.495)	d	d	d
	2	0.961 (0.838)	d	d	d

^aBased on performance of the t-test for the slope of a simple linear regression model: $Y = \alpha + \beta T + \text{error}$. Errors are normally distributed with standard deviation σ .

^bSignal-to-noise ratio = β/σ .

^cPower of two-sided t-test to detect trend, at the 0.05 (0.01) level of significance.

^dPower is approximately equal to 1 (to 3 decimal places of accuracy).

To illustrate how the numbers on the table can be applied to a real situation, consider a contaminated well with measured nitrate levels varying around an average of 800 mg/L with a standard deviation of 100 mg/L. As a result of groundwater restoration activities, nitrate levels begin to decrease at a rate of 25 mg/L per year. The signal-to-noise ratio of the well is $25/100 = 0.25$.

After 5 years of semiannual sampling, the probability is only 0.165 that a statistically significant trend will be detected at the 0.05 level of significance. The probability of statistical detection is 0.509 if the true annual decline of nitrate for the well is 50 mg/L and is 0.983 if the rate of nitrate decline is 100 mg/L per year. In practical terms, after 5 years of semiannual data collection at this hypothetical well, one can rely on a statistical test to detect a trend of 100 mg/L per year or more, but not to detect a trend of 50 mg/L per year or less.

Semiannual sampling for 10 years will reveal more subtle trends than can be observed with 5 years of data. At the 0.05 level of significance, chances remain low of detecting a real trend of 10 mg/L per year (power = 0.219); however, chances increase to probability 0.876 of detecting an average drop in nitrate of 25 mg/L per year. It is almost certain 10 years of semiannual measurements will show a statistically significant trend if average nitrate levels are decreasing in the well at an annual rate of 50 mg/L (power > 0.999).

If water sampling and analysis are done annually instead of semiannually, the power to detect a trend after a fixed length of time is lower because only half as many measurements are available for trend analysis. Compared to semiannual sampling, annual sampling will require more years to discern a trend, but somewhat less than twice as many years.

Suppose nitrate levels in our hypothetical well are slowly decreasing by 25 mg/L per year. Semiannual sampling for 10 years will produce 20 measurements. The probability is 0.876 that a trend test conducted at the 0.05 level of significance will produce a statistically significant result. If instead the water sampling had been conducted annually, only 10 measurements would be available for trend analysis. Consequently, the probability is only 0.509 that a statistically significant trend will be found after 10 years. If annual sampling is continued for another 5 years, the probability rises to 0.980 of detecting the trend. Fifteen annual measurements are more likely to show a statistically significant trend than are 20 semiannual measurements.

The information provided by statistical power calculations could be very useful for planning water monitoring programs at UMTRA sites. Table 6.1 can easily be expanded to include more or less frequent sampling and a wider range of signal-to-noise ratios than were considered in this report. To use Table 6.1, however, a value must have signal-to-noise ratio, β/σ . Historical data can provide an estimate of the denominator, σ . The numerator, β , might also be estimated from historical data if the current trend is expected to continue unchanged for several years. When power calculations are used to evaluate restoration alternatives, the slope, β , would presumably be derived from geochemical and flow models.

7.0 CONCLUSION AND RECOMMENDATIONS

Trend analysis of groundwater quality data at UMTRA Project sites is affected by non-normal errors and censored measurements. To assess groundwater trends, statistical methods must be used that are adaptable to these particularities in the UMTRA Project data base.

The most obvious evidence of non-normal errors are outliers on time-concentration plots. Outliers are isolated points that do not fit in the general pattern of the rest of the data. It is tempting to remove these unusual values from the data set; however, doing so creates the illusion that the trend is known with more certainty than may be justified.

The nonparametric trend test ranks the measurements from smallest to largest and uses only the ranks to determine whether a trend exists. Outliers are included in the analysis; however, are not allowed undue influence. The largest measurement in the data set is assigned the largest rank "n" regardless of whether that measurement is slightly or 100 times larger than all the other measurements. Nonparametric methods are extremely well suited for analyzing trends when outliers are present in the measurement data.

The use of nonparametric, rank-based procedures seems less justifiable for well-behaved data sets that have no outliers. It may seem wasteful and inefficient to throw away the measurements and use only their ranks. Surprisingly, even if the error distributions are normal, 90 percent of the useful information about the trend is retained in the ranks, so nonparametric methods are only slightly less efficient than the optimal parametric linear regression method for trend analysis under ideal conditions. Non-normality affects the performance of the estimators even when it is not so blatant as to produce outliers, and nonparametric trend methods produce reasonable evaluations of trend under a wide variety of conditions that may be present. For this reason, it is recommended that nonparametric trend analysis be used for trend analysis of UMTRA Project data.

Censored measurements are pervasive in UMTRA Project data sets. Trends in the concentration of a contaminant that take place at levels below the ability of the laboratory analytical procedure to measure will, by definition, be undetectable. It should be possible, however, to detect a change over time that raises the concentration from below detection to above or vice versa. However, the reported detection limits of many toxic constituents change over time also, generally in a downward direction. Evaluating trends in these constituents is analogous to evaluating the path of a moving target with an instrument that is also moving. The statistical methods suggested in the water-quality literature are not adequate for this task. A better approach may be discovered in the medical literature. The survival times of patients in a clinical trial are censored for a number of reasons. Patients drop out of the study, die from unrelated causes, or simply outlive the study. Innovative parametric and nonparametric procedures have recently been developed to analyze censored survival times (Kalbfleisch and Prentice, 1980). Some of these methods might be useful for analyzing censored UMTRA Project measurement data.

The Mann-Kendall trend test could also be improved for use on data sets with multiple detection limits. As mentioned in the report, the procedure censors all the data at the highest reported detection limit. In reality, although this is expedient, it is not necessary.

All the information available in censored data can be incorporated into the Mann-Kendall test statistic; however, determining its critical point would become an arduous and computer-intensive task. The UMTRA Project might consider investigating whether the necessary computer program has ever been written. If not, a professional programmer could be commissioned to produce the needed software for the Project.

In any case, the Project will need to invest in statistical software and training to maximize the potential of trend analysis methods for UMTRA Project groundwater data. Software is needed that can interface with the existing UMTRA Project data base, perform nonparametric trend analysis with or without censored measurements, and produce graphic displays of the trend. Training of UMTRA Project staff might include how human factors affect trend analysis, interpreting statistical results, and using power calculations to plan water sampling programs at UMTRA Project sites.

8.0 REFERENCES

- Cohn, T. A., 1988. "Adjusted Maximum Likelihood Estimation of the Moments of Lognormal Populations From Type 1 Censored Samples," U.S. Geological Survey, Open File Report 88-350, USGS.
- Hald, A., 1949. "Maximum Likelihood Estimation of the Parameters of a Normal Distribution Which Is Truncated at a Known Point," Skand. Aktuarietidskrift, Vol. 32, pp. 119-134.
- Helsel, D. R., 1990. "Less Than Obvious: Statistical Treatment of Data Below the Detection Limit," Environmental Science and Technology, Vol. 24, pp. 1767-1774.
- Helsel, D. R. and R. M. Hirsch, 1991. Statistical Methods in Water Resources, Elsevier Scientific Co., New York, New York.
- Hirsch, D. R., R. B. Alexander, and R. A. Smith, 1991. "Selection of Methods for the Determination and Estimation of Trends in Water Quality," Water Resources Research, Vol. 37, pp. 809-813.
- Kalbfleisch, J. D., and R. L. Prentice, 1980. The Statistical Analysis of Failure Time Data, John Wiley and Sons, Inc., New York, New York.
- Mann, H. B., 1945. "Nonparametric Tests Against Trend," Econometrica, Vol. 13, pp. 245-259.
- Sen, P. K., 1968. "Estimation of the Regression Coefficients Based on Kendall's Tau," Journal of the American Statistical Association, Vol. 63, pp. 1379-1389.
- Theil, H., 1950. "A Rank-Invariant Method of Linear and Polynomial Regression Analysis," Ned. Akd. Wentsch Proc., Vol. 53, pp. 386-392, 521-525, 1379-1412.

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

9.0 LIST OF CONTRIBUTORS

Name	Contribution
K. Smith (TAC)	Author
C. Silva (TAC)	Technical editing, document coordination
WordCenter (Vendor)	Word processing
Robert Saar	Document review
Jim Gibb	Document review
Ray Bennett	Document review

TAC Technical Assistance Contractor