APPENDIX 1
PRINCIPLES OF TREND ANALYSIS

This appendix provides information on the principles of conducting statistical trend analyses on groundwater quality data collected over an extended period of time, and on the types of statistical tests that are appropriate for this evaluation.

Appendices 2 through 7 contain graphs with nitrate concentrations plotted versus time for each of the wells evaluated. The Seasonal Kendall trend line on these graphs is hinged at median time and median concentration values. The trend line is rotated to coincide with the Sen slope.

Types of Trends

A primary goal of many water quality monitoring projects is to collect and analyze data so that changes in water quality over time (i.e., trends) can be detected. These trends can be related to both point sources and nonpoint sources; and are often related to changes in land use practices or patterns.

The two basic types of trends that can be statistically analyzed are step and monotonic. Step trends include either a sudden increase or decrease in concentration resulting from a sudden change in the primary activity controlling water quality. An example of a step trend would be a sudden increase in stream temperature downstream of a new surface water discharge. Monotonic trends are generally gradual changes that are either increasing or decreasing with no reversal of direction. An example of a monotonic trend would be the gradual decrease of groundwater nitrate concentrations as BMPs are implemented in an agricultural area.

Both step and monotonic trends can be increasing or decreasing. In addition, cycles (such as seasonal precipitation changes, tides, production schedules of industry, etc.) can be superimposed on trends. These cycles are not trends because they do not represent long-term changes.

For the purposes of this study, monotonic trend analysis techniques are believed to be most appropriate. This is largely due to the slow nature of contaminant transport in a groundwater system resulting in a relatively gradual change in groundwater quality in spite of the relatively rapid implementation of BMPs. In short, groundwater responds slowly; even to rapid changes at land surface.

Effects of Natural Fluctuations and Human Activity

It is possible for an apparent trend in water quality to be caused or masked by meteorological conditions such as precipitation cycles. It is also possible for an apparent trend in water quality to be caused or masked by human activities such as the production schedules of industry. Therefore, it is sometimes necessary to use special trending techniques to reduce the effect of outside influence (i.e., exogenous factors) on the data being examined. The purpose of adjusting the data for an exogenous variable is to reduce the background (i.e., “noise”) so that the detection of trends (i.e., “signal”) is more powerful.

For studies involving stream water quality trends, corrections are often needed to account for the flow/concentration relationship. In this study, the primary outside influence on the data is believed to be the seasonal changes in water quality caused by the irrigation season. Therefore, an evaluation of the seasonal component of water quality changes was conducted.

Measurements taken in close proximity over time are likely to be related to each other (known as autocorrelation or serial correlation), but most statistical tests require uncorrelated data (Gilbert, 1987). However, there are methods to detect serial correlation (e.g., the Durbin-Watson test). The Durbin-Watson statistic is a technique used to detect serial correlation in the residuals of a regression equation. The technique compares the residual from one time period with the residual from the previous time period, and computes a statistic that measures the significance of the correlation between these successive comparisons. The test statistic ranges from 0 to 4 and depends on the size of the data set, the number of explanatory variables in the regression equation, and the confidence level. A value near 2 indicates no serial correlation. A value near 0 indicates positive serial correlation. A value near 4 indicates negative serial correlation. There are also statistical techniques that have been developed which can account for serial correlation once it has been detected. One such technique is the Seasonal Kendall test with correction for correlation. For more information on this technique, the reader is referred to Hirsch, et al., 1984.

Loftis et.al., (1991) concludes that the distinction between serial correlation and trend is scale dependent. In other words, the distinction between serial correlation and trend is an artifact of the mathematical model used to evaluate the data as well as the time scale over which it is applied. For example, nitrate concentrations that are essentially constant over a long time (e.g., a flat trend) may contain short-term patterns which would be important from a management standpoint (e.g., decreasing trend within first half of observations). Loftis et. al., (1991) also notes that it is commonly, and probably appropriately, assumed that the scale of interest of a trend analysis is equal to the length of record (i.e., trend tests are applied to the entire record). Loftis et. al., (1991) further concludes that there is no “correct” way to approach water quality data analysis in terms of accounting for scale dependence but serial correlation can be ignored if the scale of interest is confined to the period of record.

It is clear that in order to detect or assess trends it is necessary that the data be collected at a given location using consistent collection and measurement techniques on a regular schedule and over a substantial number of years (Hirsch, et al., 1982). A change of analytical laboratories or of sampling and/or analytical procedures may occur during a long-term study. Unfortunately, this may cause a shift in the mean or in the variance of the measured values. Such shifts could be incorrectly attributed to changes in the underlying natural or man-induced processes generating the pollution (Gilbert, 1987).

Factors Complicating Trend Analysis

In order to conduct a statistically meaningful trend analysis of groundwater quality data, important assumptions regarding the data distribution (e.g., normal distribution) must be met for the chosen technique. In addition, several factors complicate the detection of groundwater quality trends. These complicating factors include seasonality, autocorrelation, missing values, outliers, and measurements near a detection limit. These complicating factors are discussed in more detail later in this report. Furthermore, results of the trend analysis must be examined for reasonableness (i.e., a “reality check”).

For example, a small but true water quality trend may not be detected in a data set with a high degree of seasonality by a technique that does not account for seasonality. As another example, if a series of measurements is reported at the detection limit, deviations from the trend line will not be normally distributed and the standard error of the least squares trend estimator will no longer apply. In many cases, outliers in the data will produce biased estimates of the least squares estimated slope itself (Gibbons, 1994).

For a steeply sloped trend, relatively few data points are necessary for the calculated values to be statistically significant. However, for a very small slope, a great deal more data may be required before the value can be confirmed as significant. Two possible consequences can occur as a result of this concept. First, two equally real trends in water quality may exist but only one will be found statistically significant because it will have a somewhat longer period of data collection. Second, an examination of an extensive data set may find a statistically significant trend that is so small as to be physically insignificant or meaningless (e.g., 0.001 mg/l/yr).

Parametric versus Nonparametric Techniques

A parametric technique is one whose validity depends upon the data being drawn from a specific known distribution (e.g., normal or log-normal). Parametric methods discussed in this report include simple least squares regression (linear regression), seasonal least squares regression, and sine/cosine seasonal least squares regression. A nonparametric (or distribution-free) technique is one whose validity does not depend upon the data being drawn from a specific distribution. The magnitude of data is ignored in favor of the relative values or ranks. Nonparametric techniques discussed in this report include the Mann-Kendall, Spearman’s rho, Seasonal Kendall without correction for correlation, and the Seasonal Kendall with correction for correlation.

If the requirements of a regression equation were known to be true (i.e., a strictly linear relationship and normally distributed residuals), then fully parametric regression would be optimal (i.e., most powerful and lowest error variance for the slope). If the actual situation departs, even to a small extent, from these assumptions then a non-parametric (i.e., Mann-Kendall) procedure will perform either as well or better (Helsel & Hirsch, 1992). If one knows that the data to be examined for trends are normal and nonseasonal, then linear regression is clearly the best. If one knows that the data are normal but seasonal, then seasonal regression may be best (depending on the magnitude of the seasonality) (Hirsch, et al., 1982).

Nonparametric procedures are always nearly as powerful as regression, and the failure to edit out or correctly transform a small percentage of outlying data will not have a substantial effect on the results (Helsel & Hirsch, 1992). The advantage of non-parametric procedures is that there are very few underlying assumptions about the structure of the data making them robust against departures from normality. In addition, the use of ranks rather than actual values makes them insensitive to outliers, moderate levels of non-detected values, and missing values.

Given that departure from normality and the presence of seasonality are common features of water quality data, coupled with the rather small loss of power associated with using the Seasonal Kendall test where the linear regression test would be most powerful, the use of the Seasonal Kendall test is recommended as an exploratory test for trend by some researchers.

Monotonic Trend Analysis Techniques

There are several types of monotonic trend analysis techniques available for use. Not all techniques are appropriate for every data set. A trend can be visually examined by plotting the observed data versus time. However, a statistical test is required to analyze the trend. If plots of the data versus time suggest a simple linear increase or decrease over time, a linear regression of the variable against time may be fit to the data. A test can be used to evaluate if the slope is different than zero. This test can be misleading if seasonal cycles are present, the data are not normally distributed, and/or the data are serially correlated (Gilbert, 1987). In fact, the results may indicate a significant slope when the true slope actually is zero (Hirsch, et al., 1982).

The Mann-Kendall test is a nonparametric procedure particularly useful in water quality evaluations since missing values are allowed and the data need not conform to any particular distribution. Also, data reported as below a detection limit can be used by assigning them a common value that is smaller than the smallest measured value in the data set. This approach is valid because the Mann-Kendall test uses only the relative magnitudes of the data rather than their measured values (Gilbert, 1987). The Mann-Kendall test analyzes the sign difference between later-measured data and earlier-measured data. Each later-measured datum is compared to all data measured earlier. An increasing trend is identified if later-measured values tend to be larger than earlier-measured values. Conversely, a decreasing trend is identified if later-measured values tend to be smaller than earlier-measured values.

If a linear trend is present, the true slope may be estimated by linear regression methods. However, the regression-calculated slope can differ greatly from the true slope if there are gross data errors or outliers in the data. Sen's slope estimator is not greatly affected by gross data error or outliers, and it can be computed when data are missing. Sen's slope estimator is closely related to the Mann-Kendall statistic in that all possible slopes are calculated between all possible data pairs and the resulting median slope is the Sen slope. The Sen's slope estimator is used to estimate the slope for the Mann-Kendall test.

If seasonal cycles are present in the data, tests for trend that remove these cycles or are not affected by them should be used (Gilbert, 1987). The seasonal least squares regression technique and the sine/cosine seasonal least squares technique remove seasonality (deseasonalize the data) while the Seasonal Kendall test accounts for seasonality in the evaluation. The Seasonal Kendall test may be used even though there are missing, tied, or non-detected values. As mentioned previously, the validity of the test does not depend on the data being normally distributed.

Hirsch, et.al, (1982) evaluated the performance of linear regression applied to deseasonalized data. This procedure (called seasonal regression) gave test results that performed well when seasonality was present, the data were normally distributed, and serial correlation was absent. However, they suggest the Seasonal Kendall test is preferred to the simple or seasonal regression tests when data are skewed, cyclic, and serially correlated. When a time series contains any non-detected values, then parametric methods of trend detection become unusable. These non-detected values present no difficulty for nonparametric methods such as the Seasonal Kendall test because nonparametric tests require making comparisons of values to determine which is the larger. The non-detected data can all be considered to be smaller than any numerical value equal to or greater than the detection limit and tied with any other non-detected value. In cases where the detection limit has changed over time as more sensitive instruments are developed, it is necessary to take all data reported below the highest detection limit (including those reported as less than any lower detection limit) and consider them all to be tied at the highest detection limit (Hirsch, et al., 1982).

A variation of Sen's slope estimator called the Seasonal Kendall slope estimator (or the Seasonal Sen Slope estimator) is used to calculate the slope for the Seasonal Kendall test. The difference is that all possible slopes within each season are calculated with the median slope being the Seasonal Kendall slope.

A variation of the Seasonal Kendall technique is also available to account for serial correlation if it is present. However, the power to detect a trend is reduced when this technique is used.

EPA (1997) recommends the following. Use the Seasonal Kendall test for hypothesis testing when testing for monotonic trends. Linear regression might also be used but is generally discouraged. If the data do not have seasonal cycles, the Mann-Kendall test could be used. The

Seasonal Kendall slope estimator is recommended when estimating the magnitude of monotonic trends when seasonality is present and the Sen slope estimator when seasonality is not present.

Table A-1 presents a comparison of seven common monotonic trend analysis techniques. Some of the assumptions regarding data distribution and technique applicability, as well as the complicating issues, are identified. Table A-1 is not intended to be a comprehensive evaluation of these techniques. Rather, it is intended to provide the reader with some basis to distinguish the techniques. Readers interested in more details on how these techniques are used in water quality evaluations are encouraged to read Gilbert (1987) and Helsel and Hirsch (1992).

Multiple Observations at Multiple Locations

When evaluating multiple sample locations with multiple observations, it may be desirable to express the results as an overall regional summary statement across all sampling locations. However, there must be consistency in behavioral characteristics across sites over time in order for a single summary statement to be valid across all sampling locations. If the stations exhibit approximately steady trends in the same direction (upward or downward), with comparable slopes, then a single summary statement across stations is valid (EPA, 2000). Gilbert (1987) stated this idea slightly differently as "when data are collected at several stations within a region or basin, there may be interest in making a basin-wide statement about trends. A general statement about the presence or absence of monotonic trends will be meaningful if the trends at all stations are in the same direction – that is, all upward or all downward."

One method of evaluating whether there is a general trend evident throughout an entire region is by performing the "Regional Kendall test" (Practical Stats Internet Site, 2000). This is done by altering the Seasonal Kendall test so that instead of testing data from all sample locations collected from a specific time interval (e.g., a particular month), data from individual sample locations collected from specific time intervals are tested. In both the Seasonal Kendall test and the Regional Kendall test, data blocks are tested individually, and then combined into one overall test result. To conduct a Regional Kendall test, blocks of data are constructed of results from a specific location during the same time period. For example, consider an example of a data set consisting of 40 wells sampled every other month for 10 years. A block of data could consist of nitrate values for a particular well sampled in January of each year (i.e., 10 data points). The test statistic is computed for each location, and then summed for all locations. The overall test statistic is divided by its standard error, a continuity correction is applied and then compared to a table of the normal distribution. The result declares whether or not there is a significant up or down trend over time for the entire region. Note that if there is an increasing trend at one location and a decreasing trend at another, they will tend to cancel one another and no overall trend may be found, even if the individual tests are significant (Practical Stats Internet Site, 2000).

Another method of evaluating whether there is a general trend evident throughout an entire region is by performing a global trend test (van Belle and Hughes, 1984). The validity of the overall trend statistic is dependent on homogeneity between seasons, between stations, and a non-significant season-station interaction term. Procedures to evaluate these criteria and evaluate a global trend are computationally intensive and are not described in this report.

LOWESS

LOWESS stands for locally weighted scatterplot smoothing (Cleveland et al., 1979). It is not a monotonic trend analysis technique. It is a data smoothing algorithm that uses a moving window superimposed over a graph of data, with analyses being performed with each move, to produce a smoothed relationship of the two variables. Data near the center of the moving window influences the smoothed value more than those farther away. The smoothed relationship is then plotted as the LOWESS line. It provides a very good graphical depiction of the underlying structure of the data. LOWESS lines are included on each of the time series plots in Appendices 2 through 7.

An advantage of LOWESS is that no model, such as a linear or quadratic function, is assumed prior to computing a smoothed line. As such, LOWESS is an exploratory tool for discerning the form of relationship between y and x. Because no model form is assumed, the data describe the pattern of dependence of y on x. LOWESS is particularly useful to emphasize the shape of the relationship between two variables on a scatterplot of moderate to large sample size.

Because a LOWESS line reflects the underlying pattern of the data and is not fitting a straight line through the data as all monotonic trend techniques do, it allows an evaluation of changes within a time series data set. For example, a monotonic trend analysis result may indicate a statistically significant downward trend in a water quality variable over a 10-year time frame. However, the LOWESS line may suggest that the water quality variable decreased for 8 years and increased during the last 2 years. As another example, a monotonic trend analysis result may not identify a statistically significant trend in a water quality variable over a 10-year time frame. However, the LOWESS line may suggest that the water quality variable increased for 5 years then decreased for 5 years. These observations might be valuable and would not be apparent from the monotonic trend analyses.

Predicting Future Concentrations

The ultimate question in analyzing time series data and computing trends is often “how long will it take?” until a particular event occurs. Answering this question requires predicting future concentrations. Predicting future concentrations with some degree of confidence requires advanced modeling techniques. This type of modeling commonly requires a considerable amount of data (e.g., hundreds of data points collected over regular intervals from a single sampling point). Environmental studies seldom include this much data. Most sample locations in this study include approximately 20 to 50 data points. Furthermore, specialized and relatively sophisticated statistical expertise is also required. Accurate prediction of future groundwater concentrations is beyond the scope of this report.

Table A-1
Comparison of Monotonic Trend Techniques
Trend Analysis of Food Processor Land Application Sites in the LUB GWMA

Trend Analysis Method	Parametric or Nonparametric	Account for Seasonality?	Advantages	Disadvantages
Simple Least Squares (Linear Regression)	Parametric	No	(1) The most powerful technique if data are normal, nonseasonal, & independent (2) Familiar technique to many people (3) Simple to compute a "best fit" line	(1) Environmental data rarely conforms to test assumptions (2) Sensitive to outliers (3) Difficult to handle non-detected values (4) Not robust against serial correlation (5) Does not account for seasonality
Mann-Kendall	Nonparametric	No	(1) Nondetects, outliers, and irregularly spaced data are permitted	(1) Does not account for seasonality (2) Not robust against serial correlation
Spearman Rho	Nonparametric	No	(1) Nondetects, outliers, and irregularly spaced data are permitted	(1) Not robust against missing observations (2) Does not account for seasonality (3) Not robust against serial correlation
Seasonal Least Squares Regression	Parametric	Yes, Deseasonalized values are obtained by subtracting monthly means averaged over years. The new values are then regressed against time.	(1) Accounts for seasonality (2) Produces a description of the seasonality pattern (i.e., seasonal means)	(1) Performs well only when data are normal (2) Not robust against serial correlation
Sine / Cosine Seasonal Least Square	Parametric	Yes, Deseasonalized values are obtained through fitting a sine curve through the data. The deviations from the curve are then regressed against time.	(1) Accounts for seasonality	(1) With few exceptions (e.g., temperature) there is little reason to believe the form of seasonality resembles a pure sine curve. (2) Performs well only when data are normal (3) Not robust against serial correlation
Seasonal Kendall without Correction for Correlation	Nonparametric	Yes, by comparing only data from the same "season".	(1) Accounts for seasonality (2) Robust against nondetects, outliers, and irregularly spaced data	(1) When applied to non-seasonal data, it has less power to detect trends than non-seasonal tests (2) Not robust against serial correlation
Seasonal Kendall with Correction for Correlation	Nonparametric	Yes, by comparing only data from the same "season".	(1) Accounts for seasonality (2) Robust against nondetects, outliers, and irregularly spaced data (3) Robust against serial correlation	(1) When applied to non-seasonal and/or non-correlated data, it has less power to detect trends than other tests.