04-RF-01164

Statistical Methods for Trending Groundwater Quality Data

Rocky Flats Environmental Technology Site

Kaiser-Hill Company, L.L.C.

Review Exemption: CEX-105-01

November 10, 2004

1/92

Review Exemption: CEX-105-01
SW-A-005020

This page intentionally left blank.

TABLE OF CONTENTS

2

3

LIST OF TABLES

4

This page intentionally left blank.

LIST OF FIGURES

5

APPENDICES
(Appendices on CD)

7

This page intentionally left blank.

EXECUTIVE SUMMARY

This evaluation has reviewed much of the published literature regarding trend analysis methods. Emphasis was placed on trend analysis of environmental water quality data. The literature review indicates a number of properties of groundwater quality data that may influence the selection of a statistical trend analysis method and its ability to correctly recognize a trend. These data properties include issues such as extreme observations (outliers), censored data, missing observations, serial correlation, and non-normal distributions.

Most of the published investigations of water quality trends since 1980 have used nonparametric statistical tests for monotonic trend. Parametric trend analysis methods require sample data to be drawn from a normally distributed population. Parametric methods that involve computation of the sample mean and standard deviation are more seriously impacted by data outliers than are nonparametric methods. Experience with groundwater quality data has also indicated that it is frequently non-normal. This shifted the focus of the report to the selection of a few candidate nonparametric methods for further evaluation.

Three nonparametric trend analysis methods that are widely used in the water quality literature were selected for further testing using groundwater quality data collected at RFETS. The candidate methods were:

- Mann-Kendall test for trend on unadjusted concentration data. The Sen's slope estimator method was used with the Mann-Kendall to estimate trend magnitudes;
- Mann-Kendall test for trend on deseasonalized concentration data. Deseasonalization was performed using the method of EPA (1989); and
- Seasonal-Kendall test for trend on unadjusted concentration data. The Seasonal-Kendall slope estimation method was used with this test.

Groundwater data for testing these methods was drawn from the "Groundwater Superset." Test data contained examples of many of the data properties listed above, e.g., outliers and missing data. Data were selected to compare trend test results for three groundwater sampling intervals (seasons) semiannual, quarterly, and monthly.

Shapiro-Wilk or Shapiro-Francia tests indicated that about 40% of the groundwater data were not normally distributed. This supported the decision to use nonparametric trend analysis methods.

The three candidate nonparametric methods gave similar trend predictions regardless of whether the input data were defined as semiannual seasons, quarterly seasons, or monthly seasons. This was surprising since much of the data was missing for monthly seasons. The methods all worked smoothly with the RFETS groundwater test data despite the presence of data issues such as outliers.

The Seasonal-Kendall test is preferred over the ordinary Mann-Kendall test for post closure monitoring at RFETS for the following reasons.

- Statistically significant seasonality was identified at 95% confidence in some of the RFETS groundwater test data. The Mann-Kendall test on unadjusted data does not account for this.

- The Seasonal-Kendall test agreed closely (88% of the time) with the Mann-Kendall test when the latter was used on deseasonalized concentration data. Thus we should use the Seasonal-Kendall test and avoid the need to deseasonalize the groundwater data.

- The Seasonal-Kendall test (used at 95% confidence) agreed well (86 to 88%) with the more subjective trend identifications made by visual inspection of seasonality plots, or of LOWESS smooths.

Statistical methods for trend testing generally assume that the concentration data are independent (i.e., not serially correlated). The Rank Von Neumann test was run to test for serial correlation in the groundwater data. Data were first deseasonalized and then detrended prior to running the Rank Von Neumann test, because it is sensitive to trend and seasonality. Statistically significant (at 95% confidence) serial correlation was found in some of the quarterly and semiannual data. The literature indicates that serial correlation is greatest at higher sampling frequencies. Therefore, semiannual groundwater sampling is preferred over quarterly or monthly sampling, for post-closure monitoring.

Measured deviations from mean water table elevation were used as an exogenous variable for trend analysis of groundwater. LOWESS smooths were used to adjust concentration data for groundwater for hypothetical impacts due to water table changes at RFETS. The Seasonal-Kendall test was applied to the adjusted data (concentration residuals) and the results were compared with those based on unadjusted data. The preliminary evidence (based on a few well-analyte combinations) indicates that water level adjustment of analyte concentrations in RFETS groundwater fails to enhance detection of concentration versus time trends.

The literature suggests that nonparametric trend analysis methods, including the Seasonal-Kendall test, are not robust against serial correlation. However, despite the serial correlation found in some RFETS test data, the trend predictions of the Seasonal-Kendall test agreed very well with the visually observed trend. It is concluded that the Seasonal-Kendall test should work well on groundwater data collected semiannually for post-closure monitoring.

A possible alternative for post closure monitoring is to use a modified version of the Seasonal-Kendall test which compensates for serial correlation (Hirsch and Slack, 1984). However, this modified test requires at least 10 years of record, and is less powerful than the ordinary Seasonal-Kendall test when the data lack serial correlation. Therefore, if statistically significant serial correlation is not found to be common in the post-closure data, the ordinary unmodified Seasonal-Kendall remains the best monotonic trend test.

ACRONYMS & TERMS

Alpha	Alpha (α) is the false positive rate or probability of making a Type I Error during statistical testing.
Alternative Hypothesis	In statistical hypothesis testing, the alternative hypothesis is accepted as true if the null hypothesis is rejected. See Null Hypothesis.
Analyte	A chemical or radionuclide whose concentration or activity in a groundwater sample is analyzed by an analytical laboratory.
ASD	Kaiser-Hill Analytical Services Division. This group establishes procedures and contracts that govern the analysis of groundwater samples collected at RFETS, and the subsequent verification and validation of the analytical data. ASD is also responsible for entering the data into SWD.
Autocorrelation	Synonym for Serial Correlation.
Beta	Beta (β) is the false negative rate or probability of making a Type II Error during statistical testing.
CDPHE	Colorado Department of Public Health and Environment.
Coefficient of Determination	The coefficient of determination (R^2) is the square of the correlation coefficient. It is a measure of the overall fit of a statistical model, such as linear regression. A perfect fit of data to a model would have an R^2 of 1.
Comparison-Wise Alpha	Maximum probability of making a false positive error for one individual statistical decision regarding one analyte in one well. See Overall Alpha.
Confidence Interval	A range of values or interval which has a know probability (or confidence) of including the true value of a population parameter (e.g., the mean).
Confidence Level	In a statistical test, the confidence level is the probability of correctly concluding that the null hypothesis is true. The confidence level equals 1 minus the false positive rate alpha.
CRDL	Contract Required Detection Limit. A synonym for RDL.
CT	Carbon tetrachloride.
D&D	Decontamination and Decommissioning.
Degrees of Freedom	Refers to the volume of data on which a statistic is based.

10

Dependent Variable	This variable (e.g., chemical concentration) may vary due to the influence of independent variables (e.g., time or distance).
Detect	A concentration at which the presence of a chemical is detected in a water sample at or greater than a reporting limit. See also nondetect.
Detection Limit	Any of several defined limits below which the concentration of an analyte cannot be reliably determined. See MDL and PQL.
DOE	United States Department of Energy.
Endogenous Variable	A dependent variable whose value is determined within a statistical model.
EPA	United States Environmental Protection Agency.
Exogenous Variable	An independent variable whose value is determined externally to a statistical model.
Explanatory Variable	A synonym for independent variable.
False Negative	In statistical testing, a false negative decision is made when the alternative hypothesis is true, but the test mistakenly fails to reject the null hypothesis. This is a Type II error.
False Positive	In statistical testing, a false positive decision is made when the null hypothesis is true, but is mistakenly rejected. This is a Type I error. The false positive rate is given by alpha.
Frequency Distribution	The absolute or relative frequencies in which concentration measurements fall into defined ranges or classes. A histogram is a graphical display of a frequency distribution. The shape of a histogram or frequency distribution may be related to a probability function, such as the normal distribution.
Hypothesis	An hypothesis is an assumption about a property of a population under study.
Hypothesis Test	This test is a statistical technique for choosing between two alternative hypotheses, the "null hypothesis" and the "alternative hypothesis." The null hypothesis is considered to be true unless there is sufficient evidence to reject it, in favor of the alternative hypothesis.
IA	The industrial area at RFETS.

Independent Variable	A measurement such as elapsed time or distance, which is hypothesized to influence a dependent variable, such as concentration.
K-H	Kaiser-Hill Company, LLC.
LOWESS	Locally-weighted scatterplot smoothing (Cleveland, 1979).
MCL	Maximum Contaminant Level.
MDL	Method detection limit.
Mean	A statistic described as the arithmetic average of a set of concentration data. The mean is the sum of the concentrations divided by the number of data points.
Median	If data are ranked or sorted in ascending order, the median is the middle value. If the number of data points is even, the average of the two middle values is the median.
Model	In statistics, a model is a mathematical description of set of data. For example, linear regression assumes that the data can be described by a linear equation.
Net Infiltration	Rainfall and snowmelt does not all reach the groundwater table. Some precipitation evaporates, runs off in streams, or is taken up by plants. Net infiltration is the fraction of water that infiltrates through the vadose zone and reaches the groundwater table.
Normal Distribution	A family of symmetrical, bell-shaped distributions whose shape is characterized by the mean and variance. The mean falls at the center of the distribution, while the spread of the data is reflected by the variance.
Nonparametric	A class of statistical tests that do not make assumptions about the shape of the underlying probability distribution, and require relatively few assumptions to be met for a valid test. Therefore, nonparametric tests may have broader applicability to environmental data than parametric tests.
Parametric	A class of statistical methods whose validity is dependent on a number of assumptions about the data. The central parametric assumption is often that the data are drawn randomly from a particular distribution, usually a normal distribution. Another common assumption of parametric tests is that the residuals (e.g., from regression analysis) are normally distributed.

12

Population	In the context of this report, a statistical population is the full set of possible analytical measurements at a monitoring well. Statistical inferences are made about the properties of this population from a small set of measurements made on water samples collected from the well.
Power	The power of a statistical test is the probability that the test will (correctly) reject the null hypothesis when the null hypothesis is, in fact, false.
ug/L	Micrograms per liter.
mg/L	Milligrams per liter.
Monotonic Trend	A type of trend involving generally smooth increases or decreases in concentration or activity over time. See also step trend.
Nondetect	Describes an analytical concentration determined to be at or below the reporting limit (RDL). The compound is either not present (i.e., a 0 concentration typically for a manmade chemical), or for ubiquitous, naturally-occurring chemicals the true concentration may be >0 but <RDL. Laboratories usually report nondetect results as the magnitude of the reporting limit with a "U" result qualifier code. Older groundwater quality data in SWD may use a "<" qualifier code. Less commonly, zero concentrations may be found in SWD for some nondetects.
Overall Alpha	The probability of a Type I error on an experiment-wide basis over all statistical tests for multiple wells or analytes. It contrasts with the comparison-wise alpha.
PCE	Tetrachloroethene.
pCi/L	picoCurie per liter.
Power	See statistical power.
PQL	Practical Quantitation Limit is a type of analytical detection limit. The PQL is the lowest concentration for which the 95% confidence interval brackets the true concentration within 20%.
QC	Quality Control, as in a QC sample generated for quality control purposes.
R	The correlation coefficient in linear regression.
RCRA	Resource Conservation and Recovery Act.

13

RDL	A Required Detection Limit specified by ASD. A synonym of CRDL.
REAL	REAL is a SWD code identifying "primary" or "real" samples, as opposed to QC samples. In this report, REAL refers to analytical data describing the primary groundwater sample collected at a well or building drain during a sampling event.
Regression Analysis	A mathematical procedure for finding the parameters of the best-fit model for the data. For example, linear regression finds the parameters (slope and y-intercept) of a linear model.
Reporting Limit	Reporting limit is often used as a synonym for detection limit. However, detection limits are often properties of an analytical method or instrument, while reporting limits are imposed on a laboratory by a client or service contract.
Residual	In regression analysis a residual is the difference between the measured value and the value predicted by the regression equation (fitted model).
RFCA	Rocky Flats Cleanup Agreement between CDPHE, DOE, and EPA, 1996.
RFETS	Rocky Flats Environmental Technology Site.
Robust Test	A statistical test which is approximately valid under a wide range of conditions (EPA, 1992).
R^2	The "coefficient of determination" is the square of the correlation coefficient and is a measure of the overall fit of a regression model. It varies from zero (no relationship) to unity (indicating that the model perfectly fits the data).
Sample Size	The number of samples or data values used to statistically describe a population.
Serial Correlation	A measure of the extent to which successive measurements or observations are related. Environmental samples collected repeatedly over short time intervals or short spatial distances, frequently show serial correlation.
SEP	The former solar evaporation ponds located in the northeast corner of the IA at RFETS.

14

Standard Deviation	The square root of the variance, or the square root of the average squared deviation from the mean of the data in a sample set. The "sample standard deviation" computes the average by dividing by n-1 rather than n, where n is the number of data points in the sample. See variance.
Statistical Power	This is the overall efficiency, strength, or ability of a statistical hypothesis test to predict a correct decision.
Step Trend	A step trend may be thought of as an abrupt change in water quality (concentration or activity) due to an event such as a contaminant spill or the implementation of a new water treatment system.
SWD	RFETS Soil Water Database maintained by ASD.
TCE	Trichloroethene.
Type I Error	See False Positive.
Type II Error	See False Negative.
Variance	The average squared deviation from the mean of the values in a sample set, or population. It is also the square of the standard deviation.
VOA	Volatile Organic Analyte.
VOC	Volatile Organic Compound, a synonym for VOA.
>=	Value on the left is greater than or equal to the value that follows the >= symbol.
<=	Value on the left is less than or equal to the value that follows the <= symbol.
>	Value on the left is greater than the value to the right of the > symbol.
<	Value on the left is less than the value to the right of the < symbol.

15

1 INTRODUCTION AND OBJECTIVES

The objectives of this report are to perform the following tasks related to trending of water quality data:

- Identify and review relevant technical literature (statistical, hydrologic, and geochemical) for potential methods of identifying water quality trends;
- Summarize the advantages and disadvantages of each potential method based on the published literature;
- Select a subset of candidate methods that appear to be suitable for trending groundwater quality data. Factors effecting method suitability, such as statistical power and statistical properties of water quality data will be defined and discussed later in this report;
- Retrieve a small set of groundwater quality data from the RFETS Soil and Water Database (SWD) for wells and analytes likely to be selected for post-closure groundwater monitoring;
- Apply the candidate methods to the set of groundwater data, evaluate their performance, and summarize their strengths and weaknesses; and
- Recommend one or more candidate methods for trending post-closure groundwater monitoring data at RFETS.

The report is organized into six sections. Section 1 states the objectives of the report. Section 2 summarizes the statistical properties of water quality data and how they may affect trend analysis. Section 3 discusses methods of trend analysis, and selects several preferred candidate methods for testing. Section 4 discusses the evaluation of candidate trending methods on test data using groundwater quality data collected at RFETS. Conclusions and recommendations are presented in Section 5. Section 6 lists the references used to develop this document. Appendices of data and statistical results are included at the end.

16

This page intentionally left blank.

2 STATISTICAL PROPERTIES OF WATER QUALITY DATA

Water quality data consist of analytical chemistry or radiochemistry measurements of the concentrations or activities of chemicals (i.e., analytes) dissolved and/or suspended in a water sample. The statistical properties of this data affect which statistical methods are most suitable for recognizing water quality trends. It is also important to decide what type of trend we are seeking to identify, as this effects the selection of a trending method.

2.1 Statistical Distribution

The statistical distribution of the population from which the samples are drawn has a strong effect on the kinds of statistical methods that are applied to the data. For their validity, parametric tests usually require certain statistical properties of a data population to be true. A common parametric assumption is that the data in the sample set were randomly selected from a normal distribution. A related condition associated with hypothesis tests for linear regression (a common trending method) requires that the residuals of the regression be normally distributed.

Professional experience has shown that there is little justification in assuming that water quality data are normally distributed. When sufficient data exist for statistical testing, water quality variables are frequently found to be non-normal (and often are positively-skewed). Hirsch and Slack (1984) report that pH, dissolved oxygen, and water temperature data are often normally distributed. However, some water quality variables fit a lognormal, or less well known distribution. For example, statistical analysis of water quality for a river in Greece indicated that river discharge, dissolved oxygen, conductivity, calcium, and nitrate concentrations followed the Weibull distribution (Antonopoulos et al., 2001). These authors also found evidence that the distribution of temperature was normal, total phosphorous was lognormal, sulfate followed the gamma distribution, and magnesium followed the logistic distribution. Hirsch and Slack (1984) also note that many types of hydrologic and water quality data are distinctly non-normal.

Professional experience has also shown that probability plots of concentration data sometimes show linear regions (of different slope) connected at inflection points. These plots suggest that the data were drawn from two distinct normal or lognormal populations. The population with the lower mean may represent background, while the other population may represent superimposed contamination.

Nonparametric statistical trending methods impose less restrictive requirements on the data and are more appropriate than parametric methods when the data are non-normal (Hirsch et al., 1991; Hirsch and Slack, 1984; Hirsch et al., 1982).

2.2 Detection Limits and Censored Data

Censoring of a statistical population of concentration data occurs when samples are not drawn from the tails of the population. This most commonly occurs when an analytical laboratory cannot measure (i.e.,

17

detect) chemical concentrations that are less than a detection limit. The statistical issue is that parametric procedures are less powerful when used on censored data, and hypothesis tests and trend slopes are less accurate (Hirsch et al., 1982). Censoring of environmental data is more of a problem for manmade chemicals (e.g., chlorinated solvent compounds) that are spatially limited in extent. Although they occur naturally, trace metal concentrations in groundwater and surface water are also frequently censored if the analytical methods employed do not have sufficient sensitivity.

Detection limits are associated with analytical methods and represent the concentration of a chemical or the activity of a radionuclide that can be quantified with a stated degree of precision. A variety of types of reporting limits have been defined in the analytical chemistry literature. Some common types are mentioned below.

Contract required detection limits (CRDLs) define the concentrations that an analytical laboratory is required to achieve to meet its contractual obligations to a client (e.g., Kaiser-Hill Analytical Services Division). The analytical methods used by the laboratory must be capable of detecting concentrations at or below the contract-specified CRDL value. In general, the term reporting limit (RL) is a synonym for CRDLs imposed on a laboratory by a client (Rong, 2002).

The instrument detection limit (IDL) reflects the analytical sensitivity of a particular instrument such as an atomic absorption spectrophotometer. The IDL is the lowest detection above the instrument noise level. A method detection limit (MDL) is associated with each analytical method. The MDL can be defined as the minimum concentration of an analyte that can be measured by the method and reported with 99% confidence that the concentration is greater than zero (Rong, 2002). More sensitive analytical methods obviously have lower MDLs. MDLs are generally higher than IDLs for the same analytes. Practical quantitation limit (PQL) is typically two to ten times higher than the MDL for an analyte.

Detection limits may vary with the chemistry of the water sample being analyzed as they are matrix specific (e.g., water, soil). Reporting limits may increase if other chemicals in the sample interfere with quantitation of the analyte (i.e., target compound). Reporting limits usually increase when it is necessary for a laboratory to dilute a sample in order to bring the analyte concentration within the linear calibration range of the analytical instrument.

Water quality data collected at RFETS frequently have multiple reporting limits for a given analyte. This has occurred because of changes in analytical laboratories under contract, changes in analytical methods used, and changes in contract-specified detection limits. Multiple reporting limits have also occurred in response to matrix effects and dilution of samples by the laboratory.

When multiple reporting limits are present in data for a single analyte, it can be reported as undetected at several different concentrations. A switch to a more sensitive analytical method with a lower detection limit will often result in detected concentrations between its low RL and older nondetects at a higher RL. This greatly complicates statistical analysis of the data and may lead to a loss of information during such analyses.

18

Nondetect data are commonly incorporated into statistical procedures by a simple substitution method. The nondetect concentration is replaced by a constant such as the full detection limit, half the detection limit, or zero. Improved estimates of the true values of nondetect data (with multiple detection limits) may be computed through advanced procedures such as maximum likelihood estimation and probability plotting procedures (Helsel and Cohn, 1988).

Some nonparametric methods can be used with data containing multiple reporting levels. However, the methods may require the analyst to recode nondetects and detects that fall below the highest detection limit, as if they were nondetects at the highest detection limit. This effectively results in a loss of concentration information from the more sensitive analytical methods.

To minimize changes in reporting limits, a post-closure Integrated Monitoring Plan (IMP) should be developed that specifies a suitable analytical method and detection limit for each contaminant of concern (COC). Consistent field sampling, sample preservation, and laboratory analysis procedures should also be followed during post-closure monitoring.

2.3 Serial Correlation Versus Independence

Serial correlation is also called autocorrelation. Statistical trend testing techniques require that the data be uncorrelated (i.e., independent). If the data are not independent, but are serially correlated, the tests will be inaccurate. Serial correlation is difficult to detect in small data subsets. It is usually detected by investigating the residuals after removing seasonality and trend (Gauthier, 2001). A common test for serial correlation is called the Durbin-Watson test (Neter et al., 1988). The Rank Von Neumann test may also be used for serial correlation in the absence of trends or cycles (IDT, 1998).

Serial correlation becomes an issue in time series water quality data when the time steps (sampling intervals) are small in comparison to the residence time of the water being sampled (Gauthier, 2001). Thus, semiannual or quarterly groundwater sampling should have less autocorrelation than monthly sampling at a given well. RFETS can minimize the effects of serial correlation by adopting a semi-annual sampling frequency for post-closure groundwater monitoring, when possible.

2.4 Seasonality and Climatic Cycles

Published literature distinguishes cyclic effects on water quality from trends. Climatic cycles tend to be irregular, multiyear changes in weather patterns. These cycles may impact local water quality mainly through changes in total annual precipitation. Climatic cycles are conventionally treated as having periods greater than one year, in contrast to seasonal cycles, which have periods up to one year (see the classical time series model in Neter et al., 1988).

Seasonal effects on groundwater quality at RFETS have been suggested but not statistically tested or verified. Although a thorough evaluation of seasonality is beyond the scope of this report, a small set of

19

RFETS groundwater quality data have been examined for evidence of seasonality. This is discussed as part of the evaluation of trend testing methods in Section 4. It is hypothesized that seasonality might effect groundwater quality through the following mechanisms:

- Net precipitation infiltration over contaminated soil areas might dissolve and transport soluble contaminants to the groundwater; or
- Net precipitation infiltration through clean soils (or insoluble contaminants) might act to dilute the existing soluble groundwater constituent concentrations.

Trend testing methods that minimize or remove seasonal effects are discussed in Section 3.

A statistical test for seasonality effects in water quality is the Kruskal-Wallis multisample test for identical populations (Bradley, 1968). Seasonality was found to be common in some surface water data collected and analyzed by the USGS (Hirsch et al., 1982).

2.5 Missing Observations Per Sampling Event

Missing observations (i.e., data) may occur by both natural and anthropogenic processes. Water samples may not be available seasonally, or during a dry year because of a lack of precipitation. Many monitoring wells at RFETS are seasonally dry and may remain dry until there is a wet year or a major storm event. Ideally for statistical analysis, these wells should be excluded from a post-closure monitoring network. Abnormally dry years could also result in an inability to sample groundwater at a well in the network. Similarly, water samples may be collected, but lost during packing and shipping, or spilled or otherwise compromised at the analytical laboratory. All of these events may result in missing observations.

Obviously, as the proportion of missing data increases, less will be known about a potential water quality trend at that monitoring location. Some statistical trending methods are less affected by missing data than others. These potential effects will be discussed in Section 3.

2.6 Multiple Observations Per Sampling Event

Multiple observations (i.e., data records) may be present in historical groundwater quality data for several reasons. Field duplicate or split (replicate) samples may have been collected and analyzed in addition to the primary water sample. After removing the field-originated QC data and the laboratory-originated QC data, historical data may still contain other sources of multiple records per analyte per well for a particular sampling event. The analytical laboratory may have diluted a sample to work within the instruments' linear calibration range or the laboratory may have re-extracted the sample and re-analyzed the extract.

Multiple observations per season may also occur when the frequency of water sampling has changed over the course of a long-term groundwater monitoring program. Presently, most wells at RFETS are sampled

20

semiannually, but RCRA wells are sampled quarterly. Based on water quality results, these routine sampling intervals may be reduced to monthly sampling for three consecutive months under requirements of the IMP. Wells that produce little groundwater are routinely visited on a monthly basis in attempts to collect water quality samples. Groundwater samples have also been collected outside the routine monitoring program to meet short-term objectives and project needs.

The problem with multiple records is that most trend analysis procedures expect a single data record per analyte per well per season. If there are multiple sampling events per season or multiple data records per season, a representative value must be selected. Statisticians refer to this process as "collapsing data". In nonparametric statistics, the median concentration may be selected, or in parametric statistics a mean may be computed to represent the seasonal concentration. Random subsampling can also be used to select one of the multiple records to represent a season. Statistically, random subsampling is considered a more effective method of collapsing data than using the mean or median (Harcum, Loftis, and Ward, 1992).

2.7 Exogenous Variables and Lag Times

Statistical trending methods may utilize exogenous variables to more clearly identify water quality trends. Lag times must be considered during trend analysis when they may delay the impact of an exogenous variable on a water quality trend.

An exogenous variable, e.g., stream discharge, is an underlying factor that may effect concentration versus time trends in surface water quality data. Potential exogenous variables for groundwater quality at RFETS are total monthly precipitation, monthly net infiltration, or variations about the mean groundwater level.

If a chemical was spilled directly into a creek via a truck accident, the local water quality impact would be almost immediate. That is, it would have a very short lag time measured in minutes. In contrast, groundwater migrates slowly, with estimated flow velocities of 41 to 717 feet per year at RFETS during 2002 (K-H, 2004a). The average flow velocity in Rocky Flats Alluvium was estimated at 91 feet per year. Therefore, VOC-contaminated groundwater leaving IHSS 118.1, for example, would require many years to migrate to the North Walnut Creek drainage, ignoring retardation effects. This is an example of a long lag time between a contaminant source and a potential impact on surface water quality.

Continuous monitoring of groundwater levels has been performed in selected wells at RFETS since 1998 as part of the real-time groundwater monitoring network (K-H, 2004a). Hydrographs from these wells are constructed with superimposed plots of precipitation data for the same time periods. These plots indicate several types of groundwater level response patterns with different lag times between precipitation event and water level change. Analysis of these responses is beyond the scope of this report, but in general, shallow alluvial wells (e.g., Wells B210489, 3686, and 6886) respond very rapidly, with lag times of less than one day. Many wells located in the IA have lag times of a few days up to 10 days. Other wells located east or southeast of the IA (e.g., Wells 1487, 05191, 03791, and 20991) appear to have water levels dependent on baseflow conditions, and their water levels show little or no response to precipitation events (i.e., they may have very long lag times).

It is believed that shallow groundwater quality at RFETS may be impacted by cyclic and seasonal variation in net precipitation infiltration or evaporation. Cyclic water quality effects may occur as precipitation infiltrates and dissolves soluble surficial or vadose zone soil contamination and transports it to the underlying groundwater. Conversely, infiltration of clean precipitation to the underlying groundwater might dilute any existing groundwater contaminants. Furthermore, evaporation of contaminated water in the vadose zone or from the shallow water table could also impose cyclic water quality effects. If these cyclic effects are significant at RFETS, they should be accounted for prior to performing the trend analyses so that underlying groundwater quality trends are recognized.

Groundwater levels are measured immediately prior to water quality sampling at RFETS and these levels likely reflect the net precipitation infiltration rate since most wells are sampled semiannually. Therefore, measured deviations from the mean groundwater elevation (or mean depth below top of casing) provide a promising exogenous variable for use in identifying groundwater quality trends. This exogenous water level variable may indicate seasonal and/or cyclic variations in groundwater quality and these water levels may be useful for discerning cyclic effects on water quality trends.

2.8 Irregularly-Spaced and Inconsistent Sampling or Analysis

Data collected for trend analysis should ideally be collected under a consistent set of sampling and analytical procedures. Changes in sampling frequency over time may impact trend testing. Often this requires interpreting the data at the lowest sampling frequency. Monthly data mixed in with semiannual data would possibly result in the exclusion of the monthly data. Irregularly spaced or convenience sampling should be avoided.

2.9 Data Outliers

Outliers are extreme observations which appear to be inconsistent with the magnitudes of the neighboring observations in time at a sampling location. Outliers may be due to data transcription errors, samples misidentified in the field or in the laboratory, or they may indicate real water quality events.

Outliers may be visually identified by inspection of concentration versus time plots. Alternatively, formal statistical tests for outliers may be run. Examples of these tests include Dixon's single outlier test, Rosner's two-tailed test for multiple, high or low outliers, or the ASTM single outlier test (Dixon, 1953; Rosner, 1975; ASTM, 1975). Rosner's test assumes a normal distribution and requires at least 25 samples. Formal testing for outliers is not suggested for post-closure monitoring. Visual identification of outliers from time series plots is sufficient for data review and evaluation. In nonparametric trending analyses, outliers do not have a strong influence on the results, unlike in parametric analyses.

22

Outlying observations are suspicious and their causes should be investigated. If errors are detected they should be corrected, or if uncorrectable, removed from the data. If the outliers cannot be shown to be due to error, then they should be retained in the data used for trend analysis. This is consistent with EPA statistical guidance on outlier retention (EPA, 1989).

2.10 Types of Trends

Two types of trends are commonly tested in water quality data - step and monotonic trends. Different statistical methods are employed to identify each type of trend. Therefore, an important consideration prior to selecting a trending method is the type of water quality trend that is expected through knowledge of events or through observation of the data. A review of the published literature indicates that monotonic trends are most frequent and most statistical trending methods apply to monotonic trends. An important point is that trends do not have to be linear. In fact, trends in water quality data are often nonlinear (Hirsch et al., 1991).

A step trend can be thought of as an abrupt change in groundwater constituent concentration due to some event such as a groundwater accelerated action (e.g., groundwater treatment system). The statistical hypothesis in testing for a step trend assumes that the data collected before a specific date (or event) are from a different population than the data collected since that date (Hirsch et al., 1991). More specifically, the test compares the pre-event and post-event means or medians.

Hirsch et al. (1991) suggest that step trend procedures should only be used for the following two cases:

1. When the concentration versus time data naturally group into two distinct time periods separated by a substantial time gap. A rule of thumb is to use a step trend procedure if the gap length is greater than 1/3 the data collection period; or

2. If a known event has occurred that is likely to have changed the water quality, then the concentration versus time record can be divided into pre-event and post-event data.

Although monitoring locations, sampling frequencies, and analyte suites have varied, RFETS has been collecting groundwater quality data relatively consistently since 1986. Therefore, substantial data gaps are not anticipated, and the 1st case probably does not apply at RFETS. However, the 2nd case may apply at RFETS and is discussed below.

A number of remediation and Site closure activities (known events) at RFETS may affect post-closure groundwater quality. Some of these known events are listed below:

- Decontamination & decommissioning (D&D) of buildings, tanks, and other structures has been on-going since 1998 and will continue into 2005. The removal of buildings built on a slab-on-grade may affect water quality when the slab is removed, through precipitation infiltration in the footprint of the former building. Building remnants, such as concrete basement walls and floors left in place, may also impact adjacent groundwater quality if the remaining portion of these structures is contaminated;

- The Water Treatment Plant (WTP) is scheduled to be shutdown during fall of 2004. Water for drinking and sanitation will no longer be imported to RFETS through ditches and pipes. This event will decrease the volume of water seeping from the ditches to the shallow groundwater and the water volume discharging to South Walnut Creek from the WTP. Thus, with less imported water, local groundwater quality will have a greater impact on gaining reaches of the creek where groundwater emerges at hillside seeps or discharges directly to the creek. The diminished volume of surface water may also effect groundwater quality beneath losing reaches of the creek;

- Removal of asphalt parking lots and roads may locally increase precipitation infiltration;

- After the buildings are removed, regrading of the land surface will potentially change patterns of surface water runoff and precipitation infiltration;

- Removal of contaminant sources such as the free-phase carbon tetrachloride at IHSS 118.1 (south of B-771) should slowly improve groundwater quality in and downgradient of that area. During 1996, there were similar source removals of VOC-contaminated soil at the T-3 and T-4 trenches located east of the Industrial Area (IA). Another source removal example is the draining and sludge removal at the Solar Evaporation Ponds (SEP) during 1995; and

- Five groundwater collection and treatment systems have been installed at RFETS to date. These systems are expected to positively impact local groundwater quality (K-H, 2002). Groundwater collection and treatment systems were installed at the following locations:

    - 881 Hillside (former OU1) in 1992 to treat VOCs and radionuclides;

    - Present Landfill seep (former OU7) from May 1996 to October 1998 a granular activated carbon (GAC) system operated to remove VOCs. Since October 1998 a passive aeration system has been used;

    - Mound Site Plume Treatment System was installed during 1997 and 1998 to remove VOCs and radionuclides;

    - East Trenches Plume Treatment System was completed during September 1999. It treats VOC-contaminated groundwater; and

    - Solar Ponds Plume Treatment System was also completed in September 1999. It treats groundwater contaminated by nitrate and uranium isotopes.

Spatially, the above remediation and closure activities may positively or negatively affect local IA groundwater quality. Temporally, these activities have occurred over many years. Because groundwater

24

migrates slowly it could take years to see the effects of these accelerated actions at downgradient wells. Different contaminant flowpaths have different lengths, hydraulic conductivities and travel times, therefore potential effects would likely be detected downgradient during different years. Groundwater treatment activities began with OU1 in 1992 followed by the other groundwater treatment activities between 1996 and 1999. Building D&D and other facility removal activities have occurred since 1998 and will extend into 2005. Site closure has no relationship to contaminant travel times; thus, no single time event is likely to delineate pre-closure effects from potential post-closure effects on a Sitewide basis.

It appears that the only way to apply step trend testing on a Sitewide basis would be to assume that 1993 through 2005 is a large data gap, transition period, or step. This approach would compare the 1986 through 1992 groundwater quality with data collected after 2005. However, it is thought that little useful information would be gained from using this approach for post-closure monitoring at RFETS.

The main objective of post-closure monitoring and trend testing is to ensure public safety and protect the environment. Another objective is to demonstrate that groundwater goals are being achieved. Monotonic trend testing should be performed as a method of confirming that these objectives are being met. Water quality trends are not always linear through time. A recent evaluation of the biodegradation of VOCs in groundwater at RFETS indicates that VOC daughter product concentrations may locally increase before eventually decreasing (K-H, 2004b).

2.11 Period of Record

Water quality data are usually collected repeatedly and systematically over a time period that spans years or decades. At Rocky Flats, this period of record for groundwater quality data for some wells extends approximately 18 years from 1986 to the present. Time series plots of groundwater quality data collected at RFETS are usually nonlinear, and may suggest upward trends during some periods and downward trends at other times. Thus, an important consideration in trend analysis is the time extent of the data to be tested for trend. Should the entire period of record be tested, or just data collected post-closure? Using all data gives the most information about historical water quality trends, but we may only be interested in recent trends at locations that monitor potential surface water impacts. This issue will be addressed in a future plan for post-closure monitoring and data assessment.

Recommendations regarding minimum number of data required for trend analysis vary in the published literature. The minimum number of data that can be used by some trending methods that evaluate seasonality effects is four data points per season. That implies collection of at least four years of groundwater data before trend analysis can start. Because of potential climatological effects the data record for trend analysis should include a representative selection of wet and dry years as well as normal water years.

This page intentionally left blank.

3 METHODS OF ASSESSING TRENDS

Trend assessment is a field of statistics and it has an extensive scientific literature. Numerous trending methods have been proposed, but not all are applicable to water quality data. This section reviews the literature and briefly describes trending methods that have been applied to water quality data.

3.1 Summary of Graphical and Statistical Trend Analysis Methods

Published literature was reviewed to determine which statistical tests have been used for detection of trends. Methods for identifying step trends are listed below with an associated reference to an author who discusses the test and/or has applied the test to water data.

- Step Trend Analysis - Parametric
    - o *Two sample t-test* (Iman and Conover, 1983).
- Step Trend Analysis - Nonparametric
    - o *Mann-Whitney* or *Wilcoxon Rank Sum test* (Bradley, 1968). This test is used with the Hodges-Lehmann estimator of trend magnitude (Hodges and Lehmann, 1963).

Many tests have been developed for monotonic trend analysis. These are bulleted below with useful references.

- Monotonic Trend Analysis - Parametric
    - o *Regression Analysis*, is usually run as linear regression for monotonic trend (Montgomery and Peck, 1982). Linear regression fits a line to the data and estimates the magnitude of a potential trend as the slope of the line. An hypothesis test (e.g., a t-test) may be run to decide if the slope is non-zero at a given level of confidence. This t-test can be misleading when the data are nonlinear, nonnormal, serially correlated, or there are seasonal or climatological effects (Hirsch at al., 1982).
    - o *Tobit estimation* (Hirsch et al., 1991). This rather complex method fits a multiple regression model to the data. The model attempts to account for time trends, season, discharge (for surface water data), and other variables. The model is fitted using the method of maximum likelihood estimation (MLE).
    - o *Combined Shewhart-CUSUM Control Charts* (Gilbert, 1987, p. 207; IDT, 1998). Control charts are a graphical tool commonly used in industrial applications of statistically-based process control. Control charts are based on parametric statistics derived from historical data. A control chart could be developed for each analyte-well

26

pair if there are sufficient detections of the analyte in the historical groundwater monitoring data. Upper control limits are established, and future concentration measurements are plotted on the chart for comparison to those limits. Control charts allow individual "out of control" measurements to be recognized whether due to an erroneous outlier, or any type of trend. They are not hypothesis tests for the presence of trend, nor do they estimate the slope of a monotonic trend.

- Monotonic Trend Analysis - Nonparametric

  - *Mann-Kendall test for trend* (Mann, 1945; Kendall, 1975). The Mann-Kendall test is functionally identical to Kendall's tau test for correlation (Kendall, 1975). The test is based on the number of times that the data increase or decrease when data values are compared to the values that follow in time. The magnitude of the trend is given by the slope estimation method of Sen (1968). This test is insensitive to outliers and can be used with irregularly spaced data or small numbers of data (Gauthier, 2001). The Mann-Kendall test can also be used on deseasonalized data. The Mann-Kendall test is affected by serial correlation and can yield incorrect results when data are strongly dependent (El-Shaarawi and Niculescu, 1992).

  - *Seasonal-Kendall test* (Hirsch et al., 1982). The Seasonal-Kendall test has the advantage of removing potential seasonal effects without trying to model their magnitude and variation. It does this by modifying Kendall's tau test to test each season separately. The individual test results are then combined into an overall test result. This test is robust against seasonality, non-normality, presence of censored data, or missing values (Hirsch and Slack, 1984). However, it is noted that this test is not robust against serial correlation. The Seasonal-Kendall Slope Estimator is an associated method of estimating the magnitude of the trend (Hirsch et al., 1982).

  - *Modified Seasonal-Kendall test* (Hirsch and Slack, 1984). This test is a modification of the Seasonal-Kendall to make it more robust against serial correlation. The modified test is more exact, but much more complex computationally. This test has also been called the Seasonal Kendall test with correction for serial correlation (Harcum et al., 1992). The modified test requires large numbers of data with at least 10 years of seasonal record.

  - *Spearman's Rank Correlation Coefficient* – Gauthier (2001) describes monotonic trend testing using Spearman's correlation coefficient (often called Spearman's rho). This test is insensitive to outliers and can be used with irregularly spaced data or small numbers of data. The Spearman test is easier to calculate than the Mann-Kendall, but this advantage is slight because of the almost universal availability of personal computers. El-Shaarawi et al. (1983) accounted for seasonality in river water by applying the Spearman method to individual monthly means.. The "Daniels test for trend" is also based on Spearman's rho (Daniels, 1950; Conover, 1999).

21

- *Cox and Stuart test for trend* (Cox and Stuart, 1955; Conover, 1999, p. 169-176). This test is a variation of the sign test and is used to detect any type of nonrandom pattern. Given a sequence of concentration data, the test pairs the later numbers with the earlier numbers and then performs a sign test on the pairs. Conover (1999, p.323) observes that tests based on Spearman's rho or Kendall's tau are considered to be more powerful than the Cox and Stuart test.

It is possible to apply the above tests to water quality data in a variety of ways. Numerous procedural variations are found in the literature.

3.2 Use of Exogenous Variables in Trend Analysis

Exogenous variables may be used to increase the power of many trend testing methods. This is based on the concept that the power and efficiency of any procedure for detecting trends and estimating their magnitude is increased if the variance of the data can be decreased (Hirsch et al., 1991). The use of exogenous variables with the Mann-Kendall test for trend is described by Alley (1988). Alley observed that exogenous variables could also be used with the Seasonal-Kendall test.

An example of this approach is based on the observation that the concentrations of major cations and anions in river water usually decrease as the river's discharge (volume of water per unit time) increases. Therefore, when seeking concentration versus time trends in surface water, investigators have often found it desirable to remove the effects of discharge (the exogenous variable) prior to concentration versus time trend testing. This can be done by regression modeling of concentration versus discharge. Temporal trend analysis is then performed on the residuals of the concentration-discharge regression. When trend testing surface water data, it has been observed that seasonality effects can remain in the data even after discharge effects have been removed (Hirsch et al., 1982).

It is important to note that if discharge versus time data show a trend, then the concentration versus discharge residuals do not necessarily indicate a concentration versus time trend (Hirsch et al., 1991). A popular modification of this technique in recent literature is to replace the regression analysis with a LOWESS smooth. LOWESS is an acronym for <u>lo</u>cally <u>we</u>ighted <u>s</u>catterplot <u>s</u>moothing (Cleveland, 1979).

It was noted in Section 2.6 that groundwater levels in a shallow unconfined aquifer directly reflect the local net infiltration rate at the point of water quality measurement. Therefore, measured deviations from the mean groundwater elevation (or mean depth below top of casing) provide a promising exogenous variable for use in identifying groundwater quality trends. This exogenous variable may indicate seasonal and/or cyclic variations in groundwater quality, and it may be useful for removing cyclic effects (both climatic and seasonal) from water quality trends at RFETS. Interestingly, use of water-level deviations to remove cyclicity has not been noted in the literature. The procedure would regress concentration versus water level deviation data. Analysis of concentration trends through time would then be performed on the residuals of the concentration-deviation regression. This analysis is performed in Section 4.7 using RFETS data.

3.3 Removal of Seasonal Variability

Seasonal effects on water quality can be minimized in several ways. In parametric trend testing, the effects of an annual cycle have been modeled by trigonometric functions of the time of year (Hirsch et al., 1991). However, it is unlikely that seasonal groundwater quality or water levels follow a smooth trigonometric function at RFETS.

EPA (1989) provides a method of removing seasonality from a set of groundwater concentration data. This method is implemented in the software program WQStat Plus (IDT, 1998). This procedure works by computing a "grand mean" concentration from all of the available data for a given analyte and well pair. Individual seasonal means are then computed for each season for the analyte and well. Finally, the individual concentration values are deseasonalized by adding the grand mean and then subtracting the appropriate seasonal mean (IDT, 1998).

The effects of seasonality could also be removed by modifying the EPA (1989) parametric seasonality correction to a nonparametric correction. A nonparametric correction for seasonality would substitute the grand median and seasonal median for their respective mean values. However, a nonparametric seasonality correction was not used in this present report as this correction is not currently implemented in the WQStat Plus software.

It is also worth noting that the water quality data are not adjusted for seasonality prior to applying the Seasonal-Kendall test. The nonparametric Seasonal-Kendall test removes the effects of seasonality without modeling it. It does this by performing the trend test on each of the seasons individually. The test statistics are then summed across all seasons. The details of this test are found in Hirsch et al. (1982).

Many of the investigators who have studied seasonality in natural waters have used monthly water sampling intervals (e.g., El-Shaarawi et al., 1983). Limited monthly groundwater quality data are available at RFETS. A small number of RCRA wells have been sampled quarterly at RFETS, while most other groundwater monitoring is performed semiannually. Thus, RFETS generally lacks the high sampling frequency data desirable for modeling seasonality. However, biannual sampling during the wet and dry seasons may exhibit seasonal effects and should be evaluated.

3.4 Trend Analysis with Censored Data

It was mentioned earlier that censored data make parametric tests less accurate than would be ideal. A parametric method of detecting trends in censored data is called "Tobit estimation" (Hirsch et al., 1991). This involves fitting a complex multiple regression model to the data that accounts for time, discharge,

29

season, and other variables. The concentration data are often log transformed to improve the fit of the model.

The nonparametric Seasonal-Kendall test and Rank Sum test may be used with censored water quality data. However, the raw concentration data must be used. The data cannot be adjusted for stream discharge or change in the groundwater table. This is because residuals cannot be computed for censored values from either regression analysis or a LOWESS smooth (Hirsch et al., 1991).

Sens' estimate of trend magnitude is associated with the Seasonal-Kendall test. The magnitude of the Sen slope will be less accurate as the percentage of censored data (nondetects) increases, however, its sign is more robust.

3.5 Trend Analysis with Serially-Correlated Data

The best way to deal with serial correlation may be to plan the water sampling program to minimize its occurrence in the data. This might be done by using semiannual or quarterly sampling, rather than higher frequency sampling such as monthly sampling. If seasonality is to be modeled, then quarterly sampling may be the best compromise. However, modeling seasonality is likely to be only of academic interest for groundwater quality data at RFETS.

El-Shaarawi and Niculescu (1992) found that the Mann-Kendall test statistic is strongly influenced by statistical dependence among the data, and the test can yield incorrect results when the dependence is strong. If autocorrelated data are suspected, the *modified* Seasonal-Kendall test can be used to correct for serial correlation (Hirsch and Slack, 1984). Note that the ordinary Seasonal-Kendall test is affected by serial correlation.

Because historic groundwater quality data at RFETS has been collected at several sampling frequencies, "data collapsing" is another method to reduce serial correlation. Data collapsing of an original time series reduces the number and frequency of data points by reducing it, for example, from monthly to quarterly or annual means. Various methods of data collapsing are discussed by Harcum et al. (1992). For example, it can be implemented by subsampling or by selecting the median value of the data that group within the newly defined sampling interval.

3.6 Selection of Candidate Trending Methods for Testing

Recommendations presented in the published literature were used to select a subset of applicable trending methods as discussed below. These candidate methods were then tested and evaluated in Section 4 using groundwater data collected at RFETS.

Step trend procedures should only be used when, 1) there is a known event that might result in a change in water quality, or 2) when the water quality data is naturally broken into two distinct periods with a long gap between them (Hirsch et al., 1991). Groundwater monitoring data have been systematically collected at RFETS since 1986, and there is no distinct data gap in the quality record since that date. As discussed earlier, groundwater remediation and closure activities have created numerous events that could hypothetically impact groundwater quality in local areas of RFETS. However, these events are spread out both spatially and temporally, and there does not appear to be a practical means of applying step trending methods on a Sitewide basis for post-closure monitoring. Therefore monotonic trend testing is recommended for post-closure monitoring, and methods of monotonic trend analysis will be considered in the remainder of this report.

Hirsch et al. (1991) published a lucid review of water quality trend testing methods. They also performed a Monte Carlo analysis that compared the performance of parametric and nonparametric methods. When there were small departures from normality, or small sample sizes, the nonparametric methods showed modest advantages in efficiency and power over parametric methods (Hirsch et al.,1991). They concluded that nonparametric procedures have only small disadvantages in power when the data are normal, modest advantages when the data depart slightly from normality, and large advantages when the data are highly non-normal. Because water quality data are frequently non-normal, they have routinely used nonparametric trending methods in their water quality investigations at the USGS.

Hirsch et al. (1982) compared monotonic trend test performance. Their results show that when data are skewed, have cyclicity, or autocorrelation, the Seasonal-Kendall test is preferred to ordinary linear regression, or to seasonal regression, and their associated t-tests.

Gilbert (1987) devotes two chapters of his widely used statistics text to water quality trending issues. Gilbert reviewed the literature on trend testing methods and concluded that nonparametric methods were the most useful for trending environmental data. Gilbert published source code for a computer program to perform nonparametric trend testing (Gilbert, 1987, Appendix B).

Gibbons (1994) also reviewed the statistical literature on water quality trend testing. Gibbons concluded that because of data issues such as outliers and nondetects, nonparametric trend analysis is most reasonable for testing groundwater quality data.

Harcum et al. (1992) observed that many investigators have moved to the use of nonparametric trend testing. However, nonparametric methods still assume independent, identically distributed error terms, and this assumption may be violated by seasonality or serial correlation. Harcum et al. (1992) compared the power of four widely used nonparametric methods while varying the properties of the test data. They varied sample size, serial correlation, fraction of missing values, seasonality, and distribution. Their recommendations for the best trend test depend on the length of the data record and the percentage of nondetects. If >10 years of monitoring data are to be trended with "no" serial correlation, and <50% missing values, they recommend use of the Mann-Kendall test on deseasonalized monthly data. For >10 years of serially correlated data they suggest the Modified Season-Kendall test. If a single test is to be selected for uncorrelated data at all record lengths they suggest the Seasonal-Kendall test. When there are

31

large proportions of missing data (>40 or 50%) they suggest using median collapsing of the sampling frequency to a longer interval.

Based on the above recommendations and the writer's experience with the non-normality of groundwater quality data, nonparametric, monotonic trend testing procedures will be selected as candidate methods for further evaluation. The nonparametric candidate trend analysis methods will include:

- Mann-Kendall test for trend on unadjusted concentration data;
- Mann-Kendall test for trend on deseasonalized concentration data; and
- Seasonal-Kendall test for trend on unadjusted concentration data.

Sen's slope estimation method will be used along with the Mann-Kendall testing to estimate the magnitude of statistically significant trends. The Seasonal-Kendall Slope Estimator (a version of Sen's method) will be used to estimate the slopes of trends identified by the Seasonal-Kendall test.

Although they are parametric, and are not candidate methods for trend analysis, Shewhart-CUSUM control charts are useful visual tools for interpreting groundwater data. Visual tools that will be applied to RFETS groundwater data include Shewhart-CUSUM control charts, seasonality charts (as implemented by WQStat Plus), LOWESS smooths, and time series plots. A number of auxiliary statistical procedures will also be used in evaluations of data properties, e.g., normality testing.

This page intentionally left blank.

4 EVALUATION OF TRENDING METHODS

Candidate trend analysis methods were selected in the previous text section. These methods will be tested to see how well they work on groundwater quality data collected at RFETS. However, software must first be selected to apply these methods.

4.1 Software Selection for Trend Analysis

Many commercially available statistical software programs are capable of performing some or all of the candidate trend analysis methods. Because normality, serial correlation, and seasonality can effect the selection of a trending method, it is desirable if the software can also perform statistical tests for these data properties. Note that software products are identified for informational purposes only, and these products are not endorsed by DOE, K-H, or URS Corporation.

WQStat Plus (copyright Intelligent Decision Technologies, Inc.) was the main statistical program selected for the present data evaluations. This program is easy to use, has convenient data deseasonalization and detrending features, and can perform Mann-Kendall, Seasonal-Kendall, and Sen's slope estimate. WQStat Plus was also used to test the data for normality (Shapiro-Wilk test), serial correlation (Rank Von Neumann test) and seasonality (Kruskal-Wallis test).

Conveniently, WQStat Plus allows the user to define up to 12 seasons and their starting dates. This feature was used to compare trend testing results for groundwater sampled semiannually, quarterly, and monthly.

A Fortran program called TREND was also used to evaluate data for this report. This code was co-developed by Gilbert and D.W. Engel and its source code is published in Gilbert (1987, Appendix B). Gilbert concluded that nonparametric methods were the most useful for trending environmental data and he developed TREND as a convenient means of applying these tests. An advantage to the reader is that in his text Gilbert interprets the results of processing four test data sets through TREND.

TREND was edited as necessary to compile under Microsoft FORTRAN, version 5. The final code modifications in version TREND3 are documented in comments within the source code. Proper functioning of TREND3 was verified by its ability to read the four test data sets of Gilbert (1987) and produce output that agrees with that discussed in Gilbert (1987). Data input to TREND3 can be exported from Excel as simple text files.

TREND3 does not have the diversity of statistical methods available in WQStat Plus. It supports the Mann-Kendall test, the Seasonal-Kendall test and Sen's slope estimates. The program can also test for homogeneity of trend among a set of different wells; for example, wells that might intersect a large VOC plume. The homogeneity feature was not used with the current data.

33

A useful freeware program for LOWESS smoothing is called Robust Fit. This code was developed by W.J. Heitler of the University of St. Andrews in Scotland. The code is based on the LOWESS methods of Cleveland (1979) and comes with example data. Robust Fit has very flexible input. For example, data can be pasted into Robust Fit from the Windows' clipboard. On output Robust Fit allows the user to copy its scatterplot graphics to the clipboard, and to save the fitted smooth as a file for further analysis.

4.2 Data Selection and Processing

The objective is to test the selected trend analysis methods on groundwater quality data representative of groundwater conditions at RFETS. This testing should reveal the strengths and shortcomings of the methods. By illuminating various data issues, the testing should assist RFETS in developing a practical statistical methodology for interpreting the post-closure water quality data.

In order to "stress test" the trending methods, analytical data were selected which included many of the data issues (e.g., multiple detection limits) described in Section 2. The following bullets describe the data selection strategy, process, and how the data were normalized.

- Wells were selected that were screened in the upper hydrostratigraphic flow system (UHSU) at RFETS. Most of the wells of interest were located in or near the Industrial Area (IA). Two wells were located near the Present Landfill, which is located north of the IA.

- Analytes were chosen that are potential post-closure groundwater COCs. Selected VOCs and nitrate/nitrite are likely post-closure COCs. Therefore, one well was chosen that monitors a plume of TCE in the IA. Another well was selected that has low concentrations of nitrate/nitrite in alluvium of North Walnut Creek.

- Field-filtered U-234 data were retrieved to test the methods' performance on radiochemistry data, which may include zero and negative activities.

- The trace, semi-metal arsenic was retrieved because it tends to have a large percentage of nondetect data, and may have multiple reporting limits. Only field-filtered arsenic data were examined.

- Data evaluated in this report were retrieved from the Groundwater Superset database, which was derived from the Soil and Water Database (SWD). Groundwater analytical results for selected analytes from each well were uploaded into a local database. Water level measurements made at the time of sampling were also uploaded.

- The full period of data collection was retrieved for each selected analyte and well, and incorporated into the working data set.

34

- Data for primary water samples (REALs) were retrieved. These data include target analyses, re-extractions, and dilutions made in the laboratory. Records rejected during validation (R) or verification (R1) were excluded from this evaluation.

- Field duplicates, equipment rinsates, and all laboratory QA/QC records were not retrieved or used in this investigation. Surrogate compounds and tentatively identified compounds were not retrieved.

- Database queries were conducted to examine the data and to identify potential problems such as incorrect concentration units, or concentration unit mismatches. When multiple concentration units were found the units, results, and detection limits were standardized to a single unit.

- Non-detect results were identified by the presence of a "U" result qualifier, and/or a "UJ" validation qualifier. Because numerous codes are found in these data fields the U values were detected in queries that use "wildcard expressions".

- Nondetect concentrations were used in two different ways. In the later discussion of method evaluations using the WQStat Plus program, nondetects were used at face value (i.e., at the reporting limit). WQStat used Cohen's adjustment to means and standard deviations to compensate for nondetects. In the later evaluations using TREND3 nondetects were used at one half the reporting limit.

- Formal statistical testing for outliers was not performed, although the data were examined visually for potential outliers.

- An objective of this investigation was to compare the performance of trend testing methods on seasons of different lengths. For example, the Seasonal-Kendall test is described in the literature on the basis of 12 monthly seasons. At RFETS we wish to test it using 2 semiannual seasons (for post-closure monitoring), and with quarterly seasons (for RCRA groundwater monitoring).

- In creating the test data, seasons were defined on a calendar basis for simplicity, rather than trying to match the precipitation cycle at RFETS. Quarterly seasons start on January 1st, April 1st, July 1st, and October 1st. Monthly seasons start on the 1st day of each month. Semiannual seasons begin on January 1st and July 1st.

- The master set of retrieved groundwater quality data was exported from Access as an Excel spreadsheet and is found in Appendix A.

- In order to create seasons of different lengths, the Appendix A spreadsheet was copied to three daughter spreadsheets in Appendices B, C, and D. These daughter spreadsheets were edited as follows to create semiannual, quarterly, and monthly data.

- Appendix A often contains multiple data records per sampling event for a given well-analyte combination. Reasons for this were discussed in Section 2.6. For trend analysis a single data record is desired per combination of well, analyte, calendar year of sampling, and season. After sorting each daughter spreadsheet by well, analyte, and sample date, numbered seasons were defined. Semiannual seasons were defined in the data of Appendix B. Quarterly seasons were defined in the data of Appendix C, and Monthly seasons were defined in Appendix D.

- Pseudo-random subsampling of the daughter spreadsheets in Appendix B, C, and D, was performed by manually selecting one record to represent each season, when multiple records were present. This was done without the bias of looking at the reported concentration, result qualifier, or validation fields. Pseudo-random sampling was considered sufficient for creation of test data, but more rigorous random sampling designs should be employed if subsampling is used during post-closure monitoring.

- Working data for trend analysis were exported from the Excel spreadsheets of Appendices B, C, and D. The spreadsheets were exported in various file formats to be readable by the statistical software. WQStat Plus, for example, has a data file translator that reads tab delimited text files containing the relevant data fields.

4.3 LOWESS Smooths of RFETS Groundwater Data

Program Robust Fit (Heitler, 2004) was used to plot the analyte concentration versus time data of the RFETS groundwater data of Appendix A. Robust Fit also generated a LOWESS smooth or curve to assist the visual identification of trends and to help identify statistical issues (data problems). The 19 plots created by Robust Fit follow later in this section.

LOWESS was developed by Cleveland (1979). Inspection of the plots is intended to show how challenging trend detection can be in real-world data. The widely varied shapes of the plots indicate that trends are frequently nonlinear, and tendencies may reverse direction one or more times over the period of record.

Other properties of the data that may cause difficulties for trend detection methods and software are noted here. Outliers are seen in some of the plots (e.g., Figure 4-11 nitrate/nitrite in Well 70193). Periods of missing data are noted in the following plots: Figure 4-16 arsenic in Well B206989 (8 years missing), Figure 4-18 PCE in Well P114889 (5 years missing), and Figure 4-13 VC in Well P115689 (5 years missing). Many of the data sets in Appendix A contain multiple data values per season, regardless of whether a season is defined as a month, quarter, or half-year. Multiple data points in a season are a problem for some statistical methods and computer programs. Multiple data points originate through changes in sampling frequency (e.g., quarterly to special monthly sampling), or when a laboratory performs a dilution or re-extraction of a sample. Some trend testing software can use the mean or median of the multiple records in a given season.

34

Another problem is that the 1.7 year period of record for U-234 in Well P415889 (Figure 4-19) is probably too short for meaningful trend testing. This is because: 1) groundwater moves slowly and its chemistry tends to change slowly, 2) less than three or four years is considered too short to characterize seasonal effects, and 3) multiyear climatic effects have not yet been adequately determined.

Robust Fit (and some trend testing software) cannot discriminate between detect and nondetect data. Therefore, the plots were made by including nondetect concentrations at one-half the detection limit. Input data are listed in Appendix A.

The smoothing algorithm used either 1st order or 2nd order polynomial smoothing as noted in the caption below each plot. It is important to note that in LOWESS, a 1st order smooth usually does not yield a straight line through the data. LOWESS is not a simple linear regression line fit. Second-order smoothing is usually more rounded than 1st order smoothing.

The caption notes the number of data values used in the plot as "n." Finally, the "half-window" is a LOWESS fitting parameter. It represents the number of points on either side of the point being smoothed, which are included in the polynomial fit. Large values of the half-window give smoother plots at the expense of loss of small detail.

User control over axis labeling is very limited in the plotted output of Robust Fit. The vertical axis always denotes concentration in µg/L for non-radionuclides, or activity in pCi/L for radionuclides. The horizontal axis of these plots is always elapsed time in decimal years since the start of water quality monitoring for a given analyte - well combination.

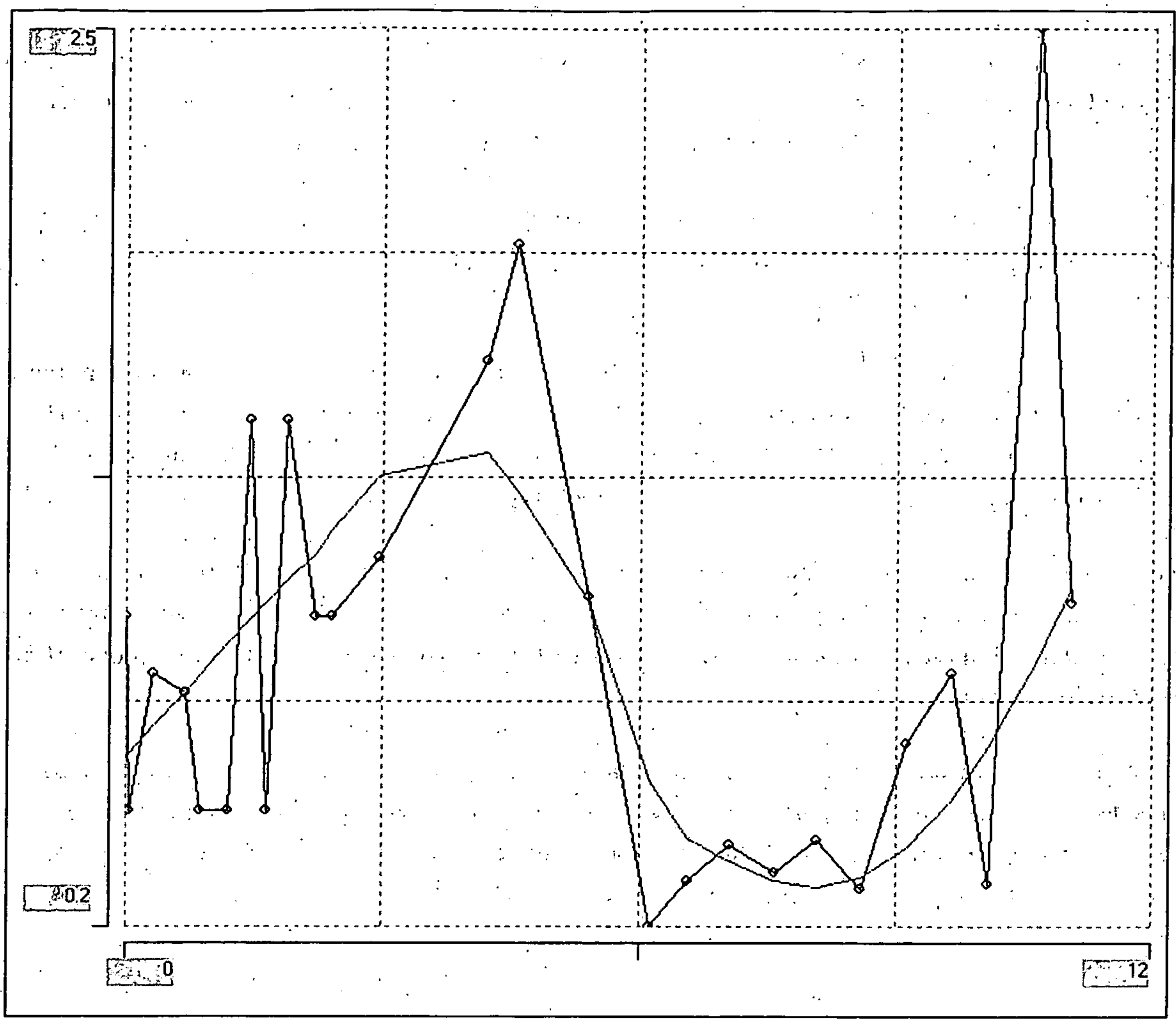

Figure 4-1 Smooth of Dissolved Arsenic Concentrations in Groundwater at Well 03991.

Smooth performed over 12 years of record (N = 27 data points, 2[nd] order smooth, half-window = 9). Although a nonlinear trend or cycle in implied, 89% of the data are nondetect and the range of the vertical concentration axis is only 2.3 µg/L. Therefore, it's not clear if there is a real cycle or trend above the sampling and analytical variability. Well 03991 is located in the northeast trenches area south of Pond B1.

38

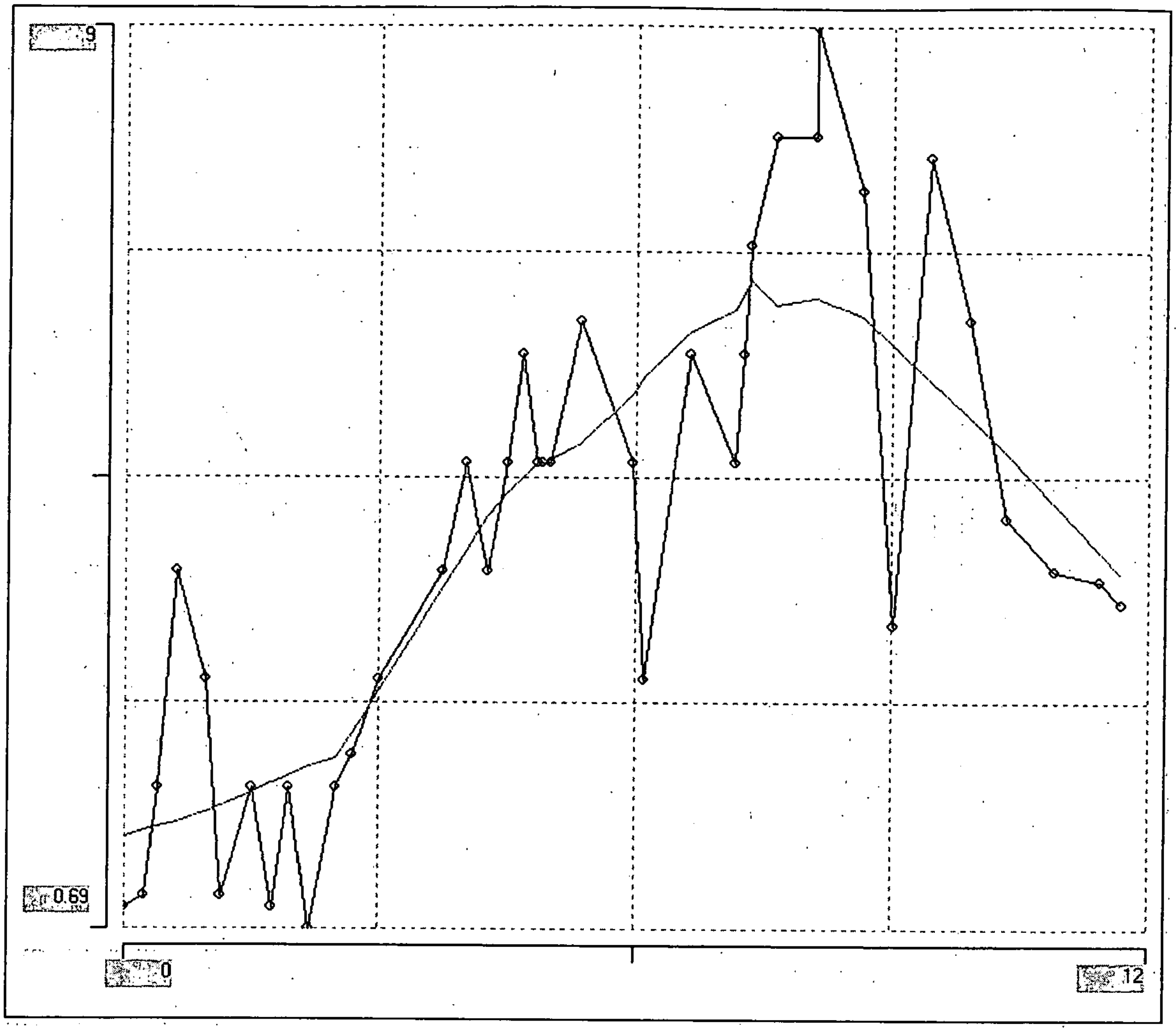

Figure 4-2 Smooth of Carbon Tetrachloride Concentrations in Groundwater at Well 06091

Smooth performed over 12 years of record (N = 39, 1st order, half-window = 9). Both the data and the smooth indicate that CT concentrations exhibited an increasing trend for the first 8 years, then started to attenuate. Well 06091 is located in the East Trenches area.

39

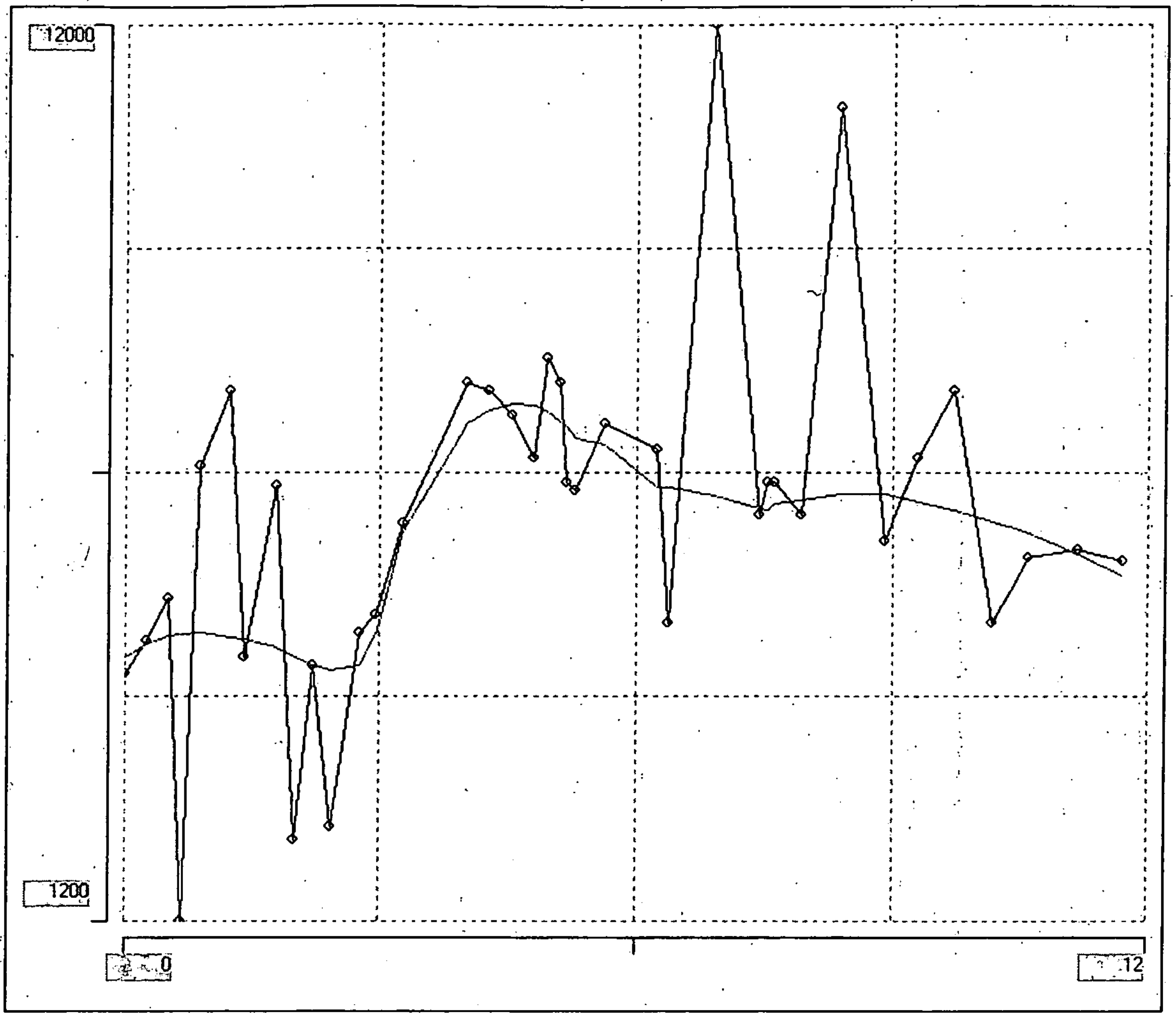

Figure 4-3 Smooth of Nitrate/Nitrite (as N) Concentrations in Groundwater at Well 06091

Smooth performed over 12 years of record (N = 38, 2nd order, half-window = 8). Well 06091 is located in the East Trenches area.

HO

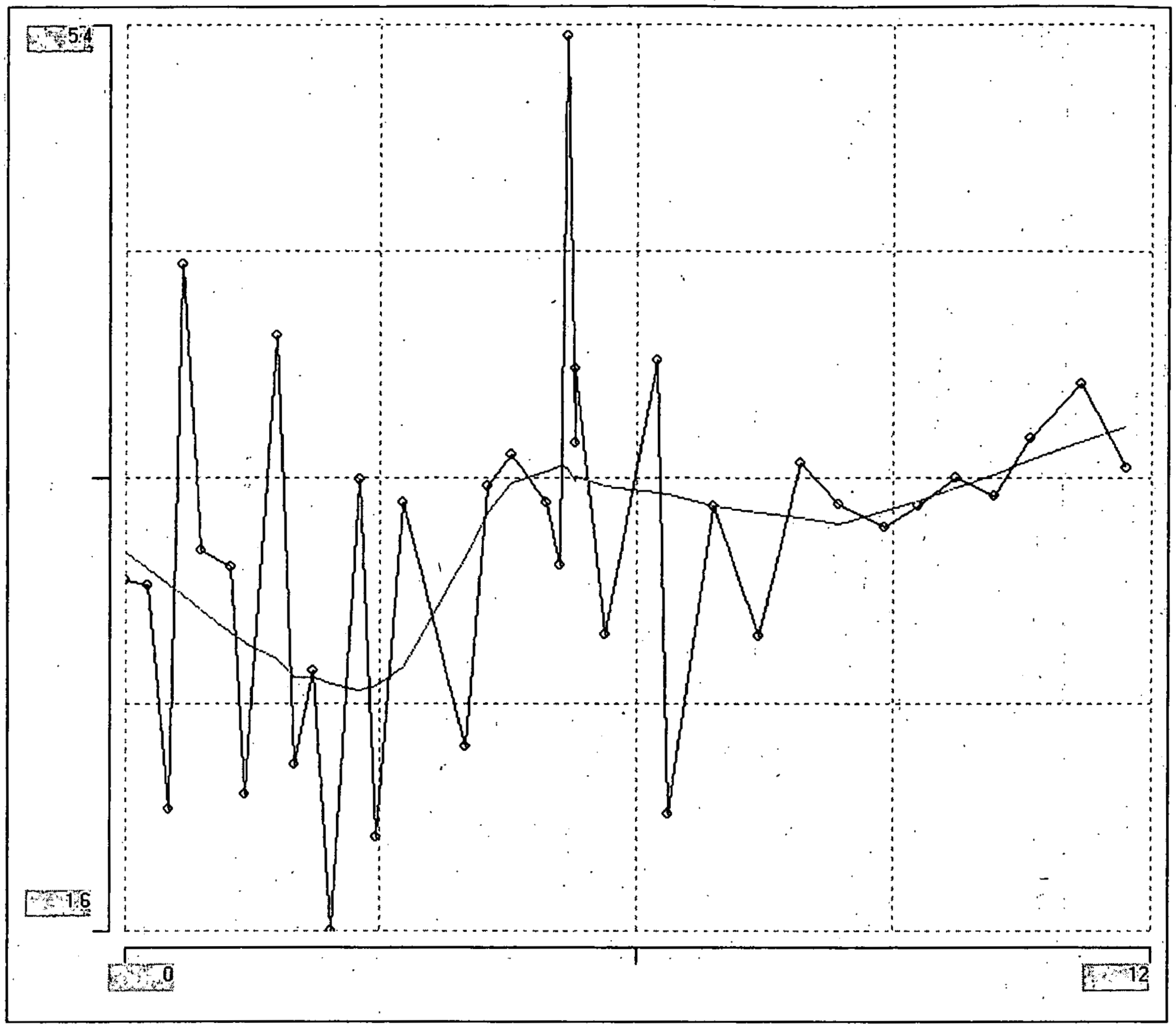

Figure 4-4 Smooth of Uranium-234 Activity in Groundwater at Well 06091

Smooth performed over 12 years of record (N = 36, 2nd order, half-window = 9). Visually it is difficult to decide if there is an upward trend. Well 06091 is located in the East Trenches area.

41

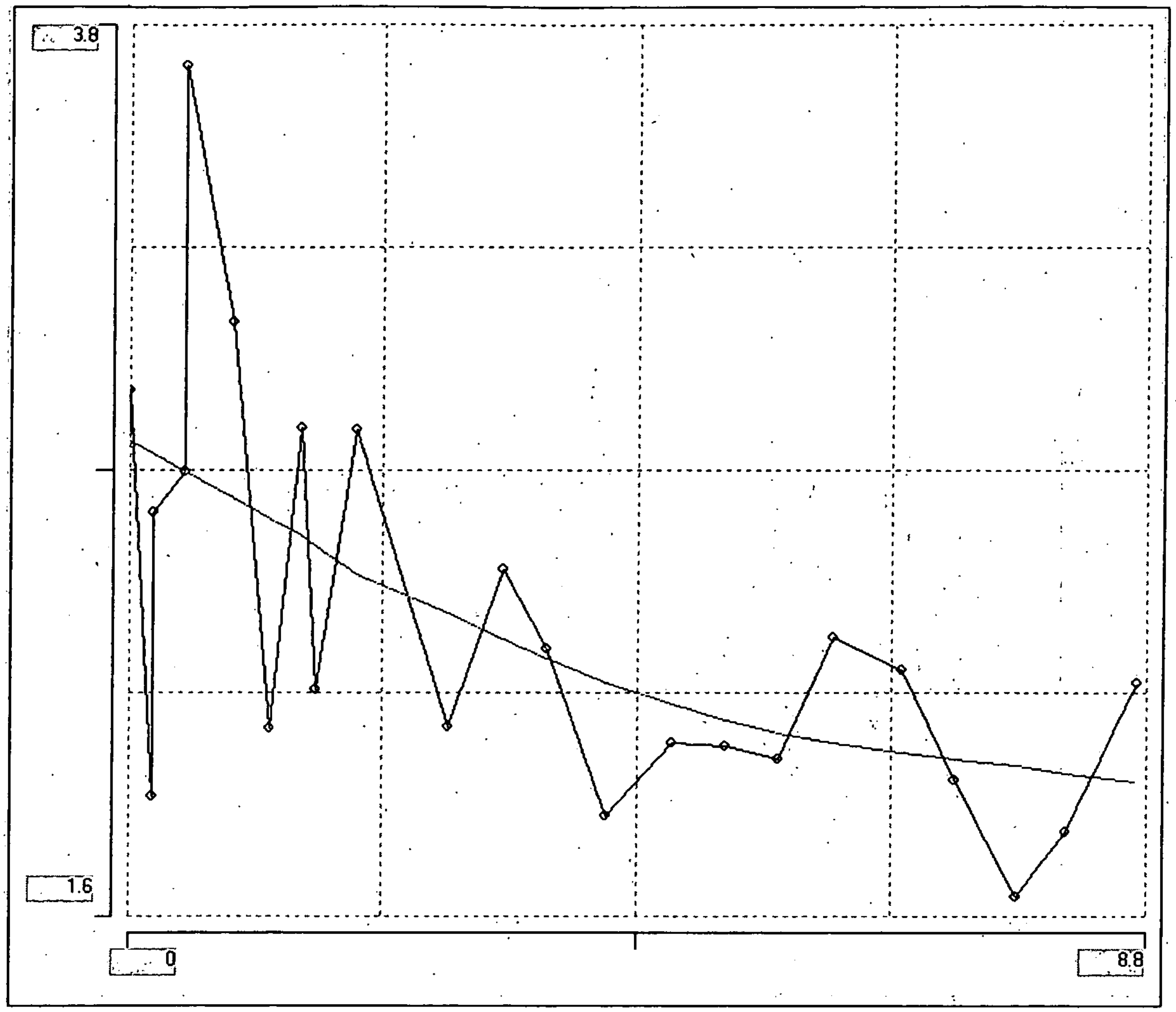

Figure 4-5 Smooth of Uranium-234 Activity in Groundwater at Well 10194

Smooth performed over 9 years of record (N = 24, 1st order, half-window = 8) and shows a decreasing trend. Well 10194 is located in the East Trenches area.

H2

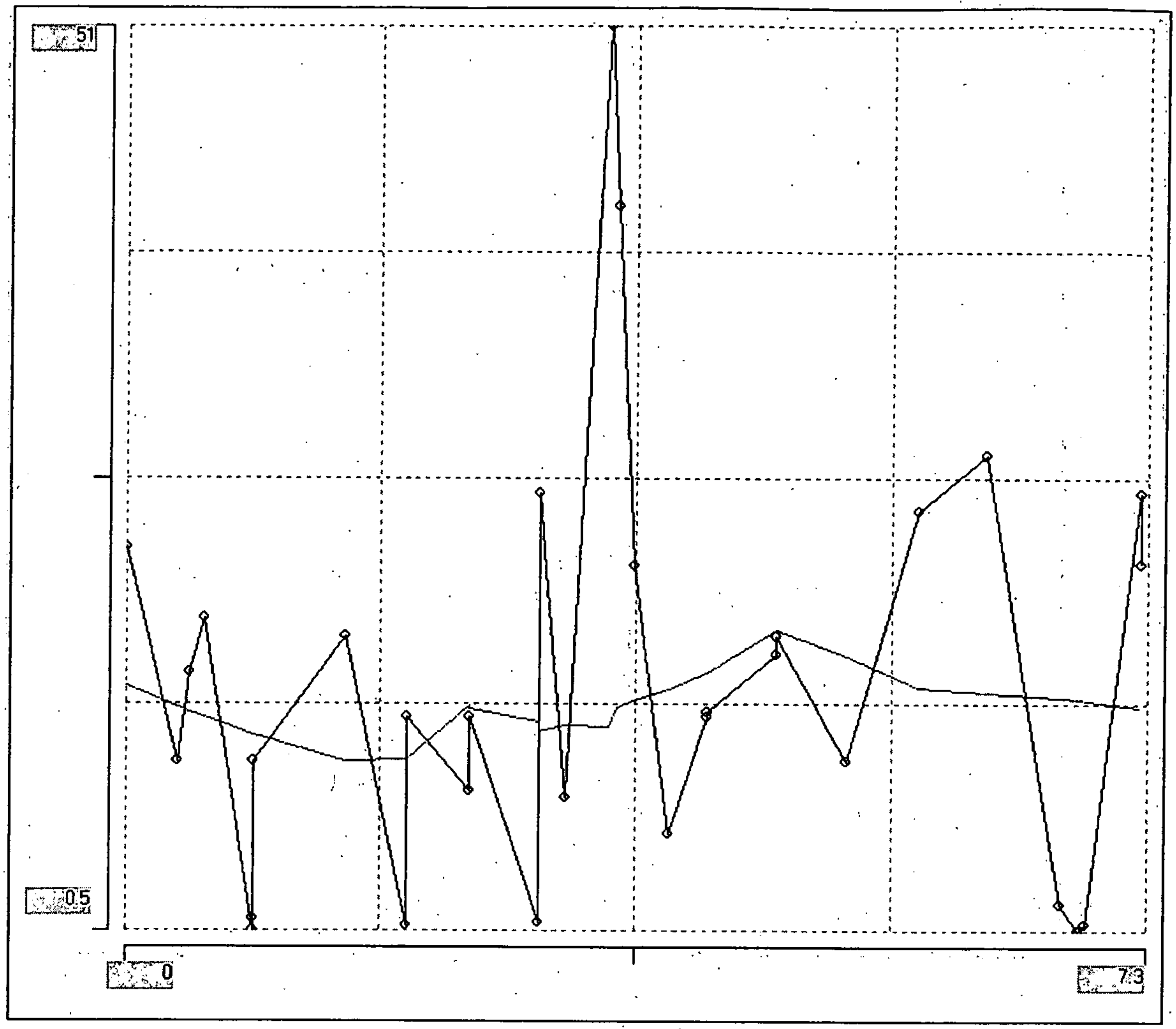

Figure 4-6 Smooth of Chloroform Concentrations in Groundwater at Well 23296

Smooth performed over 7 years of record (N = 31, 1st order, half-window = 9). The three data points in the middle of the plot appear to be a real increase in CF rather than outliers. About 23% of the data are nondetect. Well 23296 is located in South Walnut Creek and west of Pond B3.

43

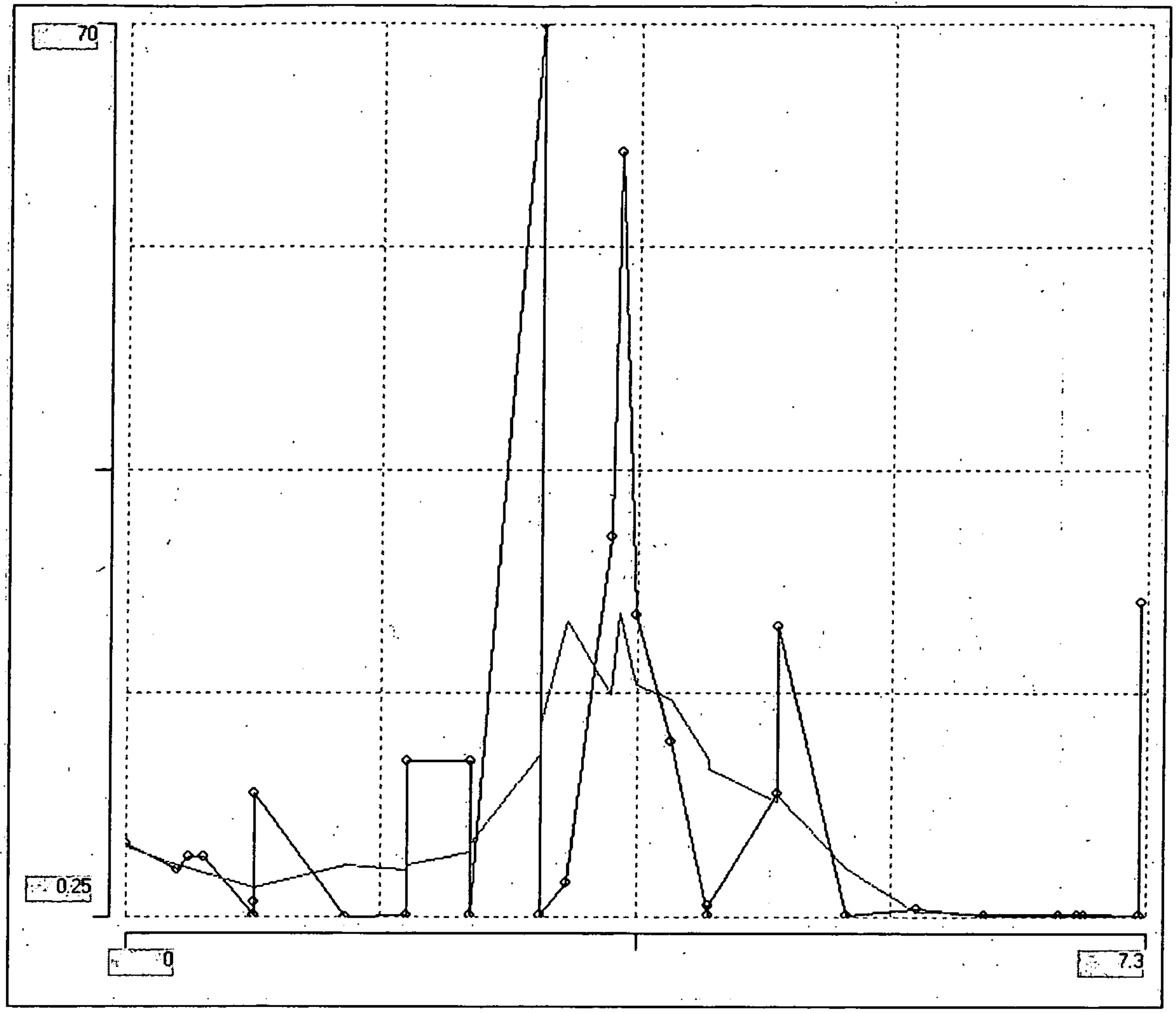

Figure 4-7 Smooth of Methylene Chloride Concentrations in Groundwater at Well 23296

Smooth performed over 7 years of record (N = 31, 1st order, half-window = 6). Similar to the previous CF smooth, the data points near the middle of this plot indicate real increases in MC concentration. Actually, 74% of the data are nondetect with some elevated detections. Well 23296 is located in South Walnut Creek and west of Pond B3.

44

Review Exemption: CEX-105-01

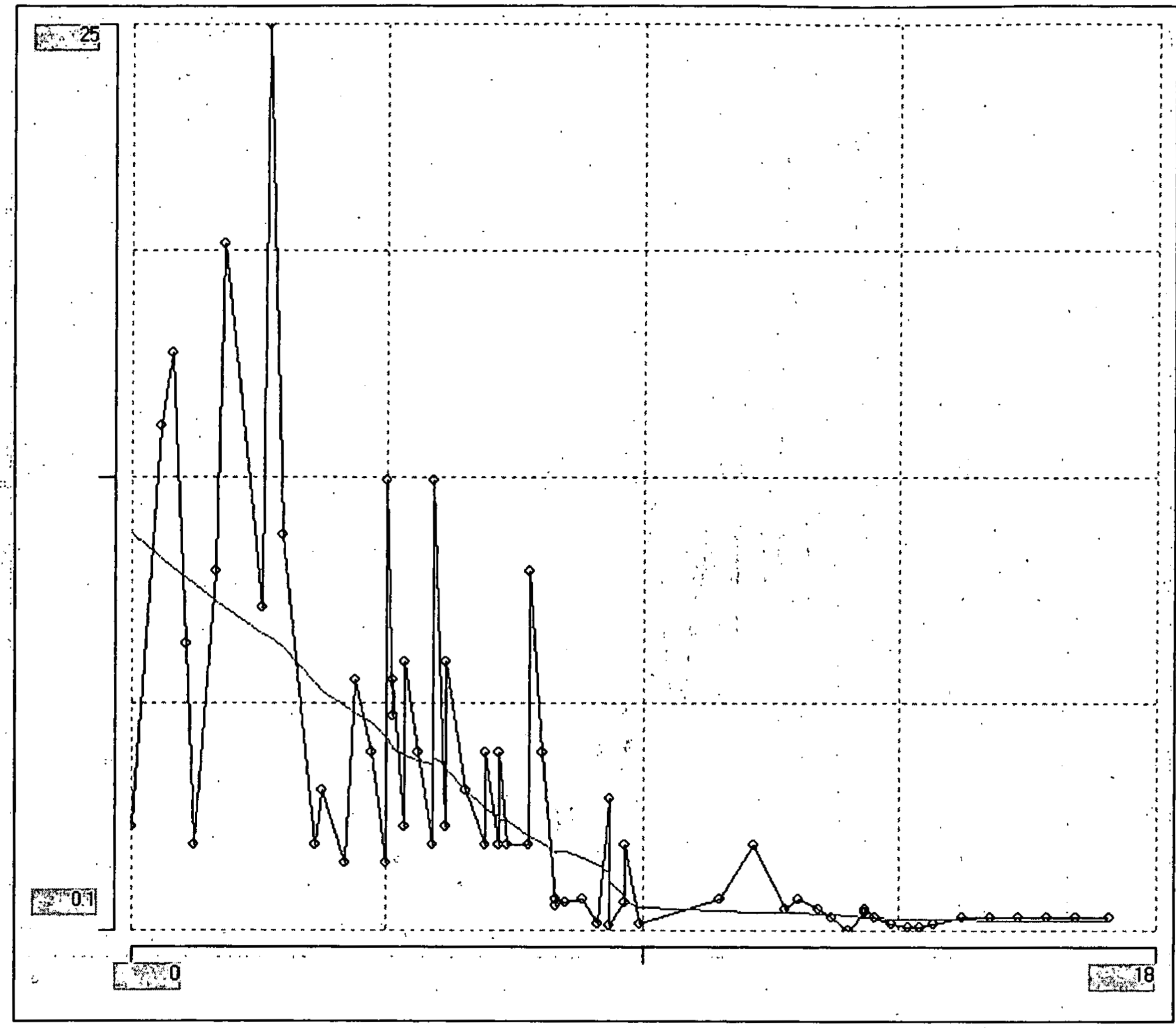

Figure 4-8 Smooth of Trichloroethene Concentrations in Groundwater at Well 3586

Smooth performed over 9 years of record (N = 68, 1st order, half-window = 14) and shows the attenuation of TCE to the detection limit. Forty seven percent of the data are nondetects, mostly at the lower right of the plot. Well 3586 is located just north of the Mound Site plume groundwater treatment system, which was installed during 1997 – 1998. The midpoint of the x-axis is 1994 so the decline occurred prior to the treatment system.

45

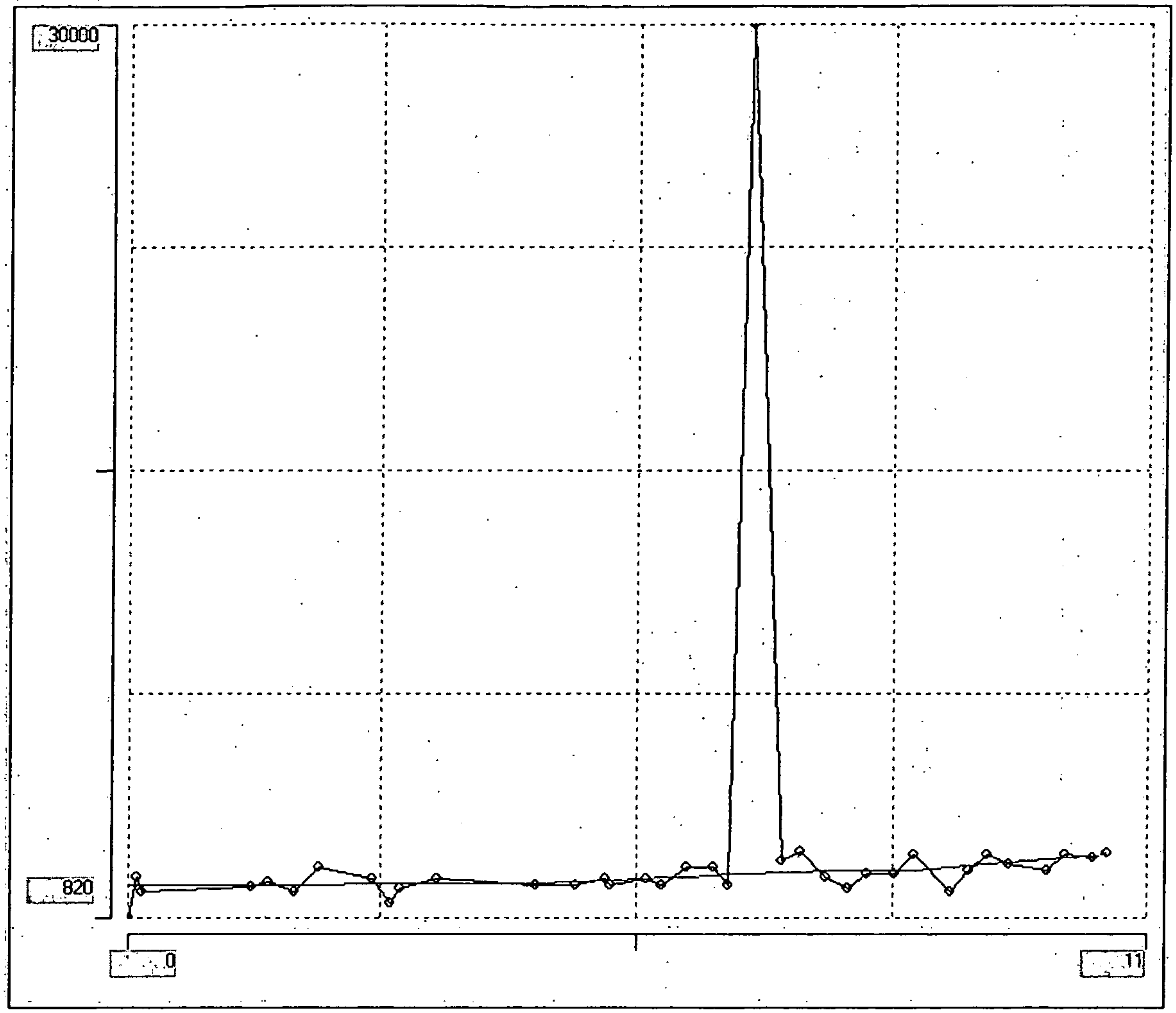

Figure 4-11 Smooth of Nitrate/Nitrite (as N) Concentrations in Groundwater at Well 70193

Smooth performed over 11 years of record shows a single large outlier superimposed on a slight upward trend (N = 36, 1st order, half-window = 8). Note how the smooth ignores the outlier. Well 70193 is a RCRA (upgradient) groundwater monitoring well at the Present Landfill.

48

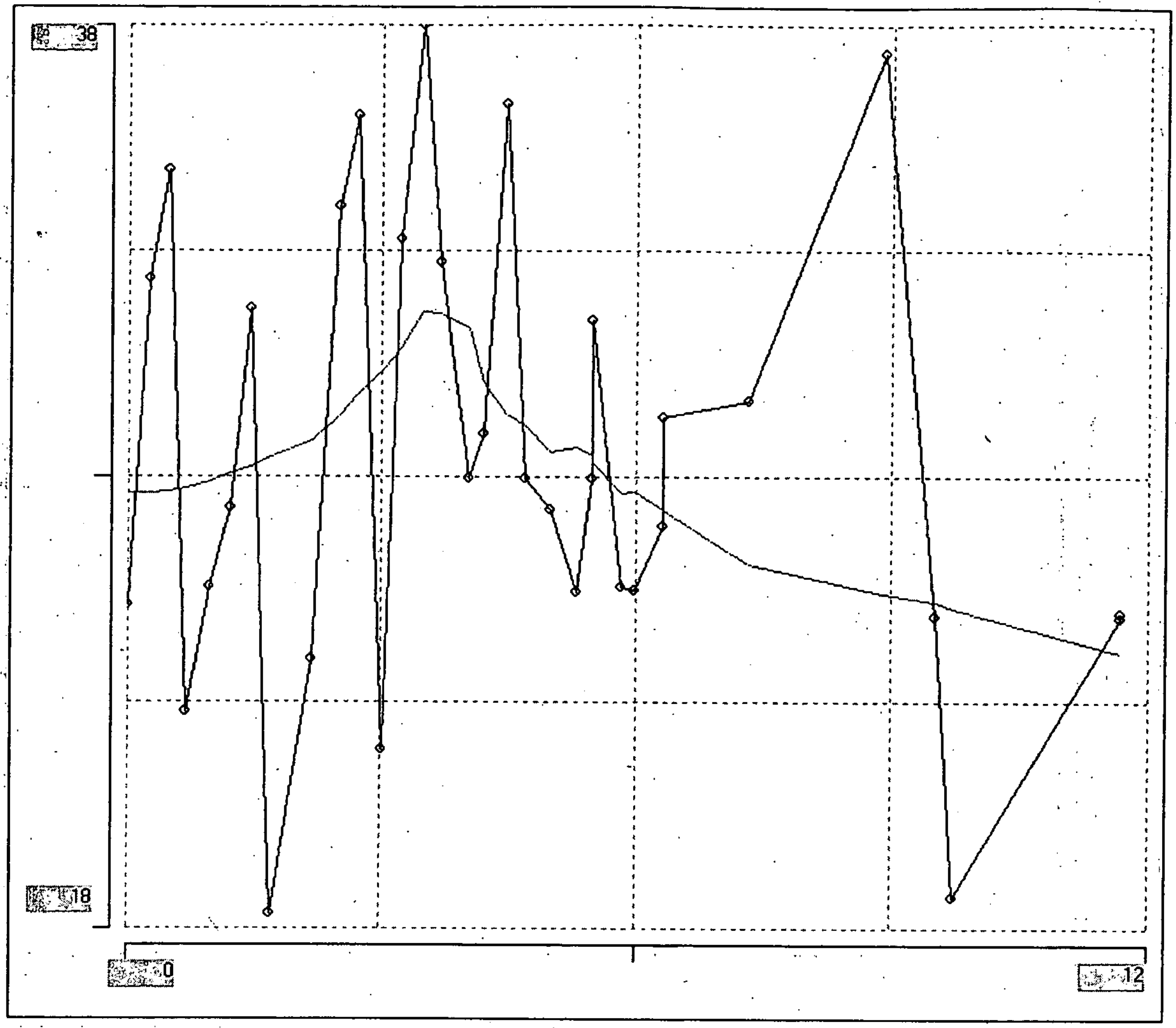

Figure 4-12 Smooth of Uranium-234 Activity in Groundwater at Well B210489

Smooth performed over 12 years of record (N = 33, 1st order, half-window = 8). First and 2nd order smooths of this data suggest an activity maximum exists after about five years of monitoring. In the author's opinion, that maximum is not noticeable to the naked eye if the smooth were absent. B210489 is located east of the former Solar Evaporation Ponds (SEPs).

49

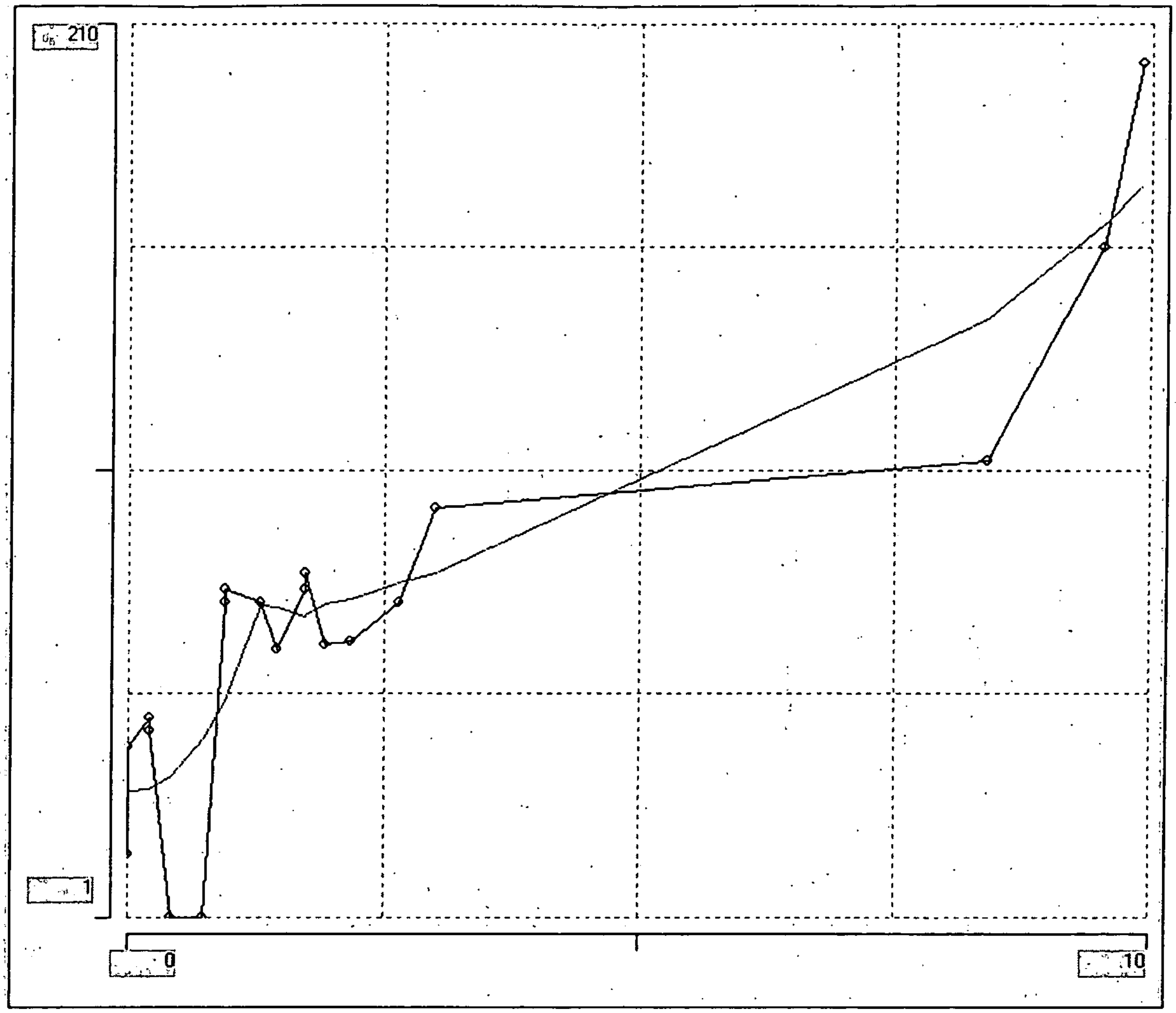

Figure 4-13 Smooth of Vinyl Chloride Concentrations in Groundwater at Well P115689.

Smooth performed over 10 years of record indicates an increasing trend in VC concentration (N = 19, 1st order, half-window = 4). These data contain about 10% nondetects. This well is located east of B551.

50

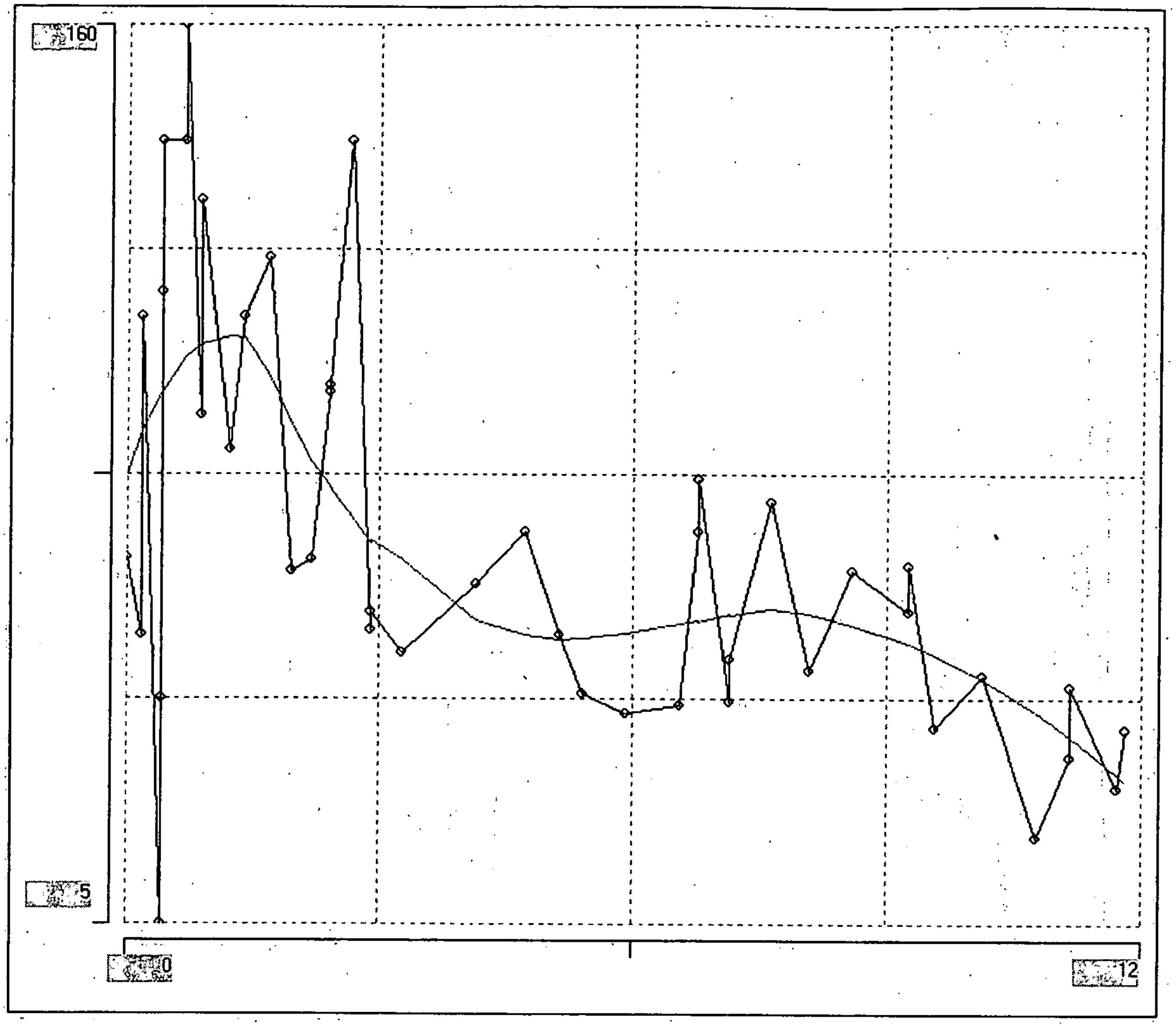

Figure 4-14 Smooth of Trichloroethene Concentrations in Groundwater at Well 00491

Smooth performed over 12 years of record (N = 44, 2nd order, half-window = 12). Various 1st and 2nd order smooths of this data all indicate the TCE concentrations peaked after one or 2 years of monitoring, followed by a nonlinear downward trend during the next decade. Nondetects are only 2% of this data. Well 00491 is located in the americium zone.

51

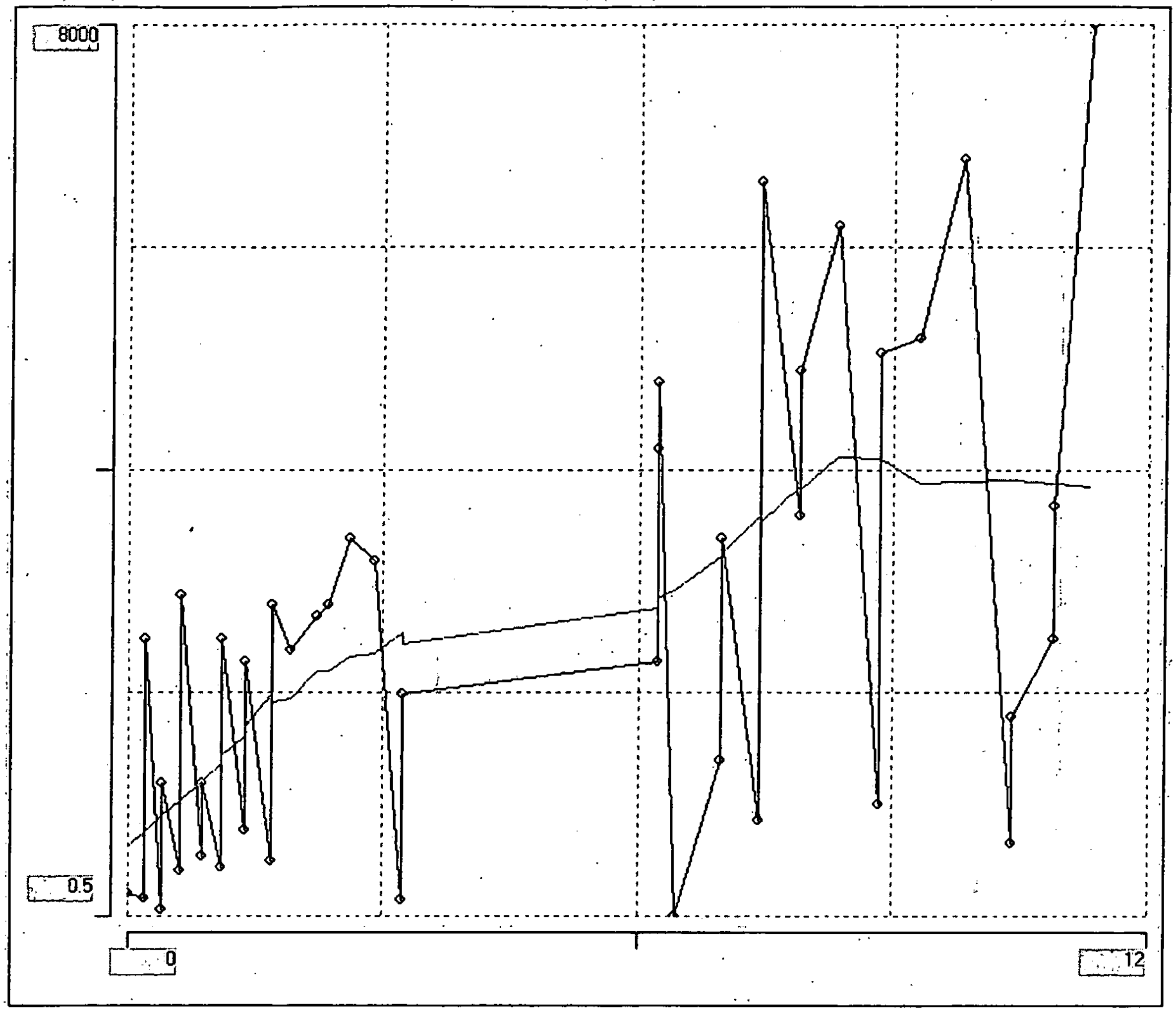

Figure 4-15 Smooth of Tetrachloroethene Concentrations in Groundwater at Well 02291

Smooth performed over 12 years of record (N = 42, 1st order, half-window = 8). Despite the variability in PCE concentrations, both the smooth and the raw data imply the presence of an increasing concentration trend. These data contains only 2.4% nondetects. The well is located in IHSS 113 at the Mound Site.

52

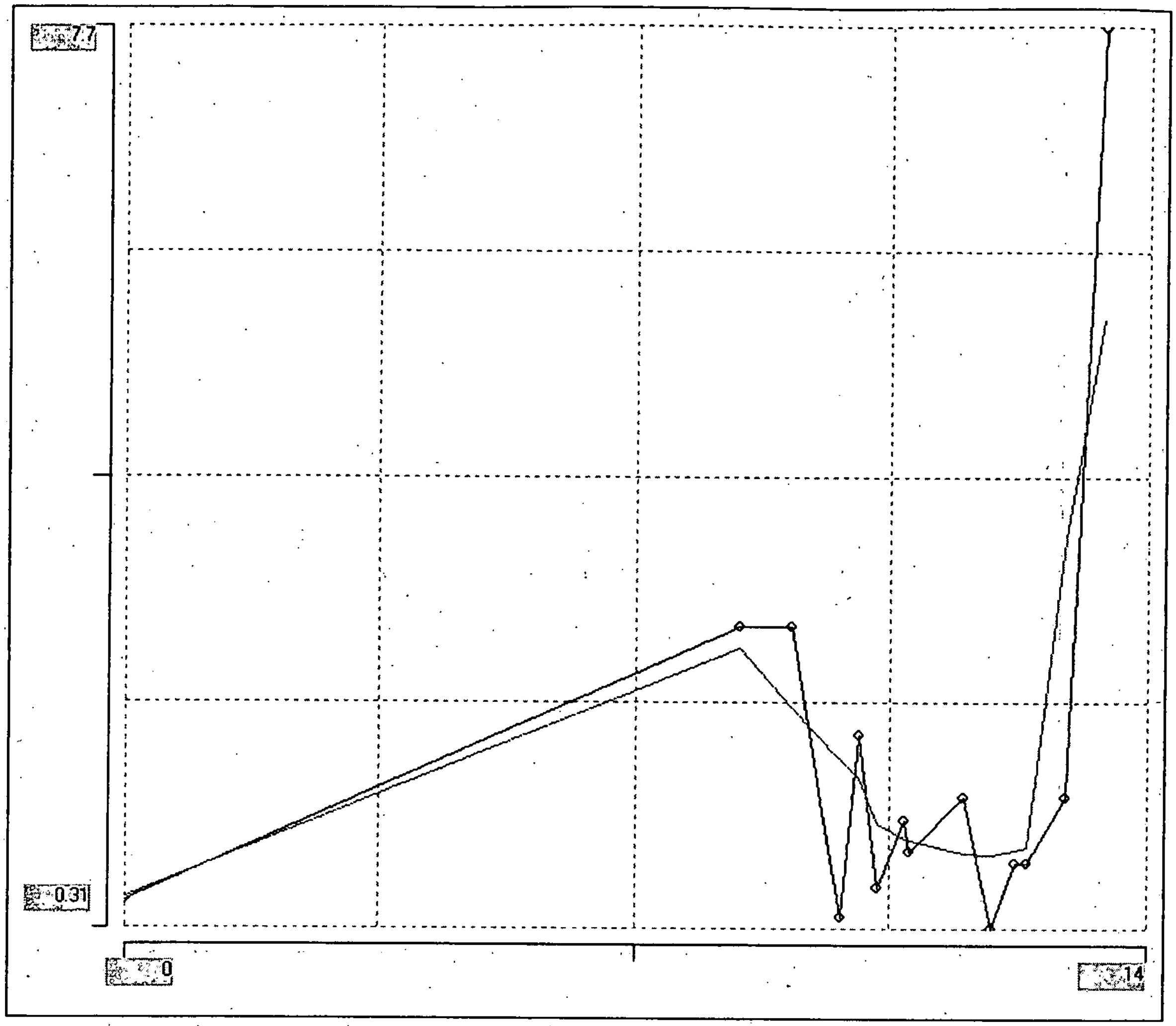

Figure 4-16 Smooth of Dissolved Arsenic Concentrations in Groundwater at Well B206989

Smooth performed over nearly 14 years of record (N = 14, 1st order, half-window = 4). It may not be defensible to evaluate a trend in this data because it consists of 50% nondetects. This well is located in No Name Gulch east of the East Landfill Pond.

53

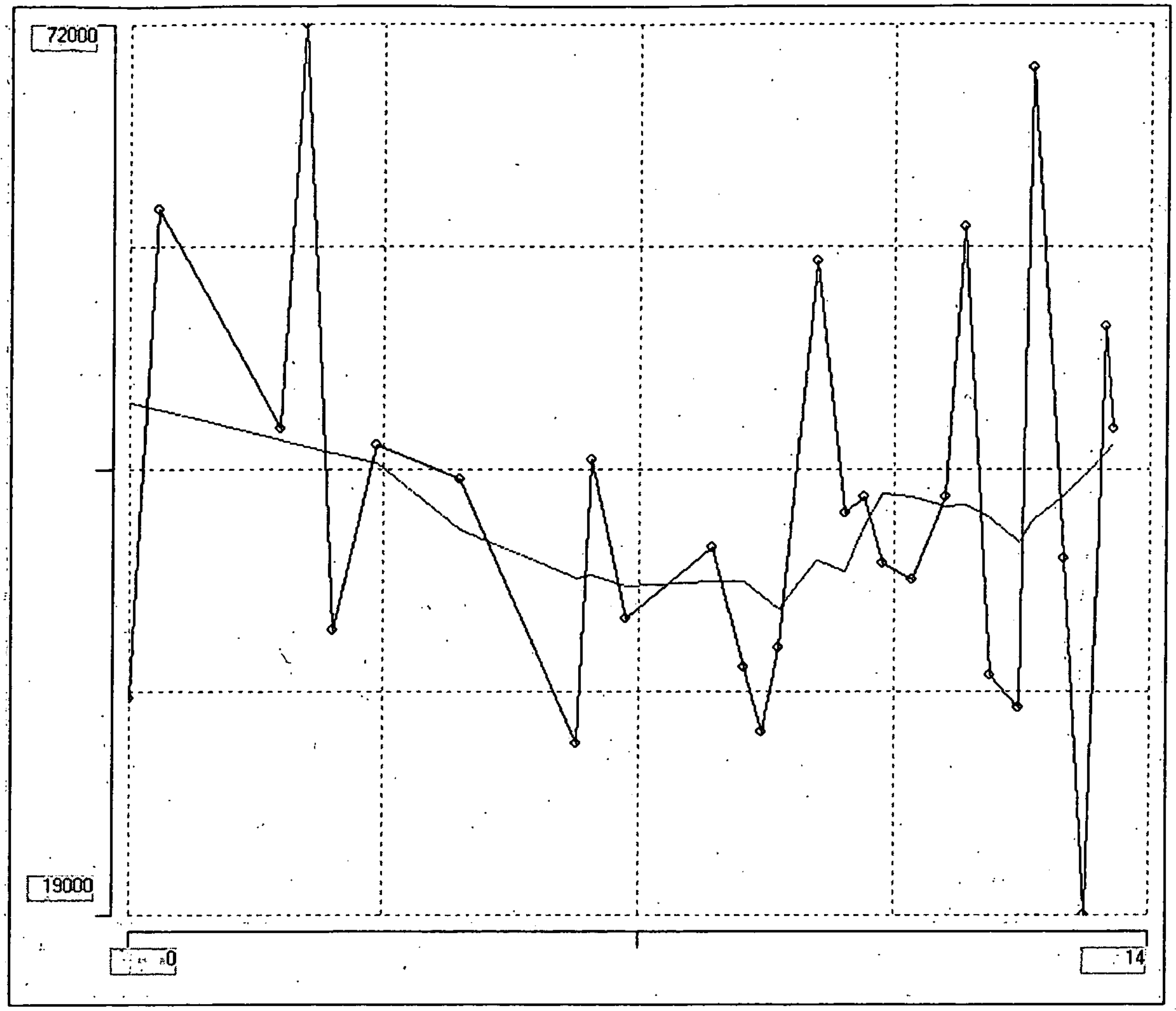

Figure 4-17 Smooth of Nitrate/Nitrite (as N) Concentrations in Groundwater at Well B206989

Smooth performed over 14 years record (N = 28; 1st order, half-window = 6). Various 1st and 2nd order smooths fitted to this data imply a shallow concentration minimum in the data at 6 to 9 years of record. Overall, there does not appear to be an increasing or decreasing trend. There are no nondetects in the data. This well is located in No Name Gulch east of the East Landfill Pond, and its groundwater contains nitrate/nitrite concentrations greater than the 10 mg/L surface water action level.

54

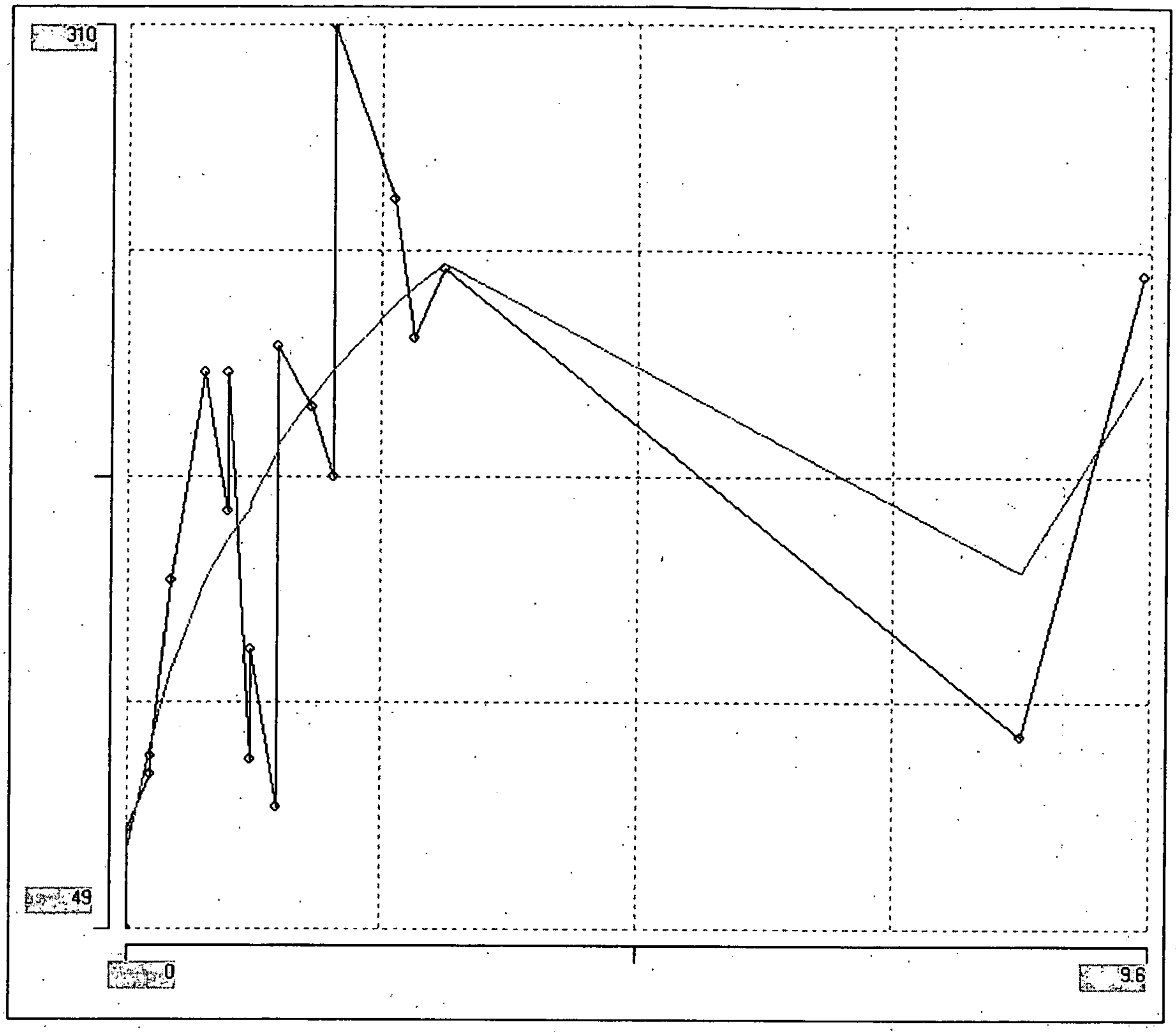

Figure 4-18 Smooth of Tetrachloroethene Concentrations in Groundwater at Well P114889

Smooth performed over 10 years of record (N = 20, 2nd order, half-window = 8). There appears to have been a steeply rising trend for the first three years, then a large gap in monitoring until year 8.4. Step-trend testing methods might be applicable to this data if 1) a hypothesis is proposed that explains why the initial trend stopped and 2) there were more than two recent data points. This well is located south of B371.

55

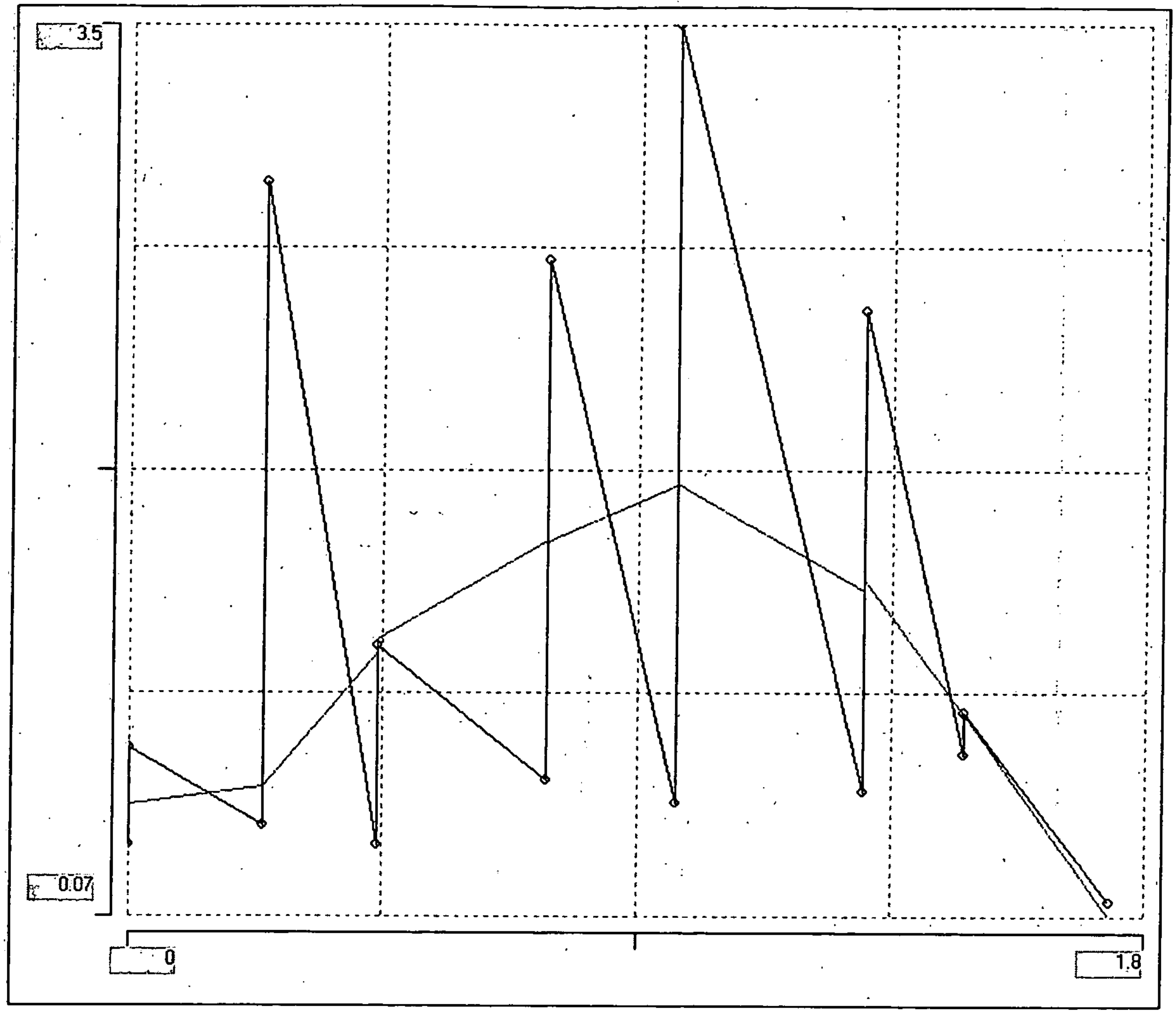

Figure 4-19 Smooth of Uranium-234 Activity in Groundwater at Well P415889

Smooth performed over 2 year record (N = 15, 1st order, half-window = 3). Less than four years of record is probably too short of a period for trending. The well is located west of B119.

56

4.4 Statistical Evaluations of RFETS Data Using WQStat Plus

The RFETS groundwater data described earlier were evaluated in several ways to support trend testing decisions. It is important to remember that these data evaluations are for statistical testing purposes, and they are not intended to describe overall groundwater quality at RFETS. A much more complete and much larger data set would be required for the latter purpose.

4.4.1 Plots of RFETS Data

Test groundwater data from RFETS were initially interpreted based on visual inspection of three types of concentration versus time plots generated by WQStat Plus,

1. Time series plots by analyte for each well.
2. Seasonality plots showing deseasonalized concentrations versus time along with the original data
3. Shewhart-CUSUM charts by well-analyte pair.

These plots are contained in Appendix E. They were used to subjectively estimate if an upward trend, downward trend, or no trend at all was present in the data. Strong trends with steep slopes that are approximately linear are easy to identify visually or statistically. However, trend identification is more subjective when the data plot shows minima or maxima, is nearly flat-lying, or contains a trend reversal.

When data from multiple wells were plotted on the WQStat Plus time series plots (stacked time series), the concentration scales were often compressed making it difficult to visually identify trends. The Shewhart-CUSUM charts for individual wells were often easier to interpret.

Seasonality plots are best for visual interpretation of trends, because they plot the deseasonalized data versus time. This effectively lowers the noise level and emphasizes any trend in the data. Where possible, seasonality plots were used in this report to visually identify trends. Later tables will compare these visual estimates of trend with the statistical trend test results.

Note that three parameters must be specified when plotting a Shewhart-CUSUM chart. These are: "h" a control limit for the CUSUM value, an upper Shewhart control limit "SCL", and "c" the acceptable displacement of the standardized mean. WQStat default parameters were used in the present work, because control limits for post-closure monitoring have not been defined. Therefore, the reader is cautioned that while the Shewhart-CUSUM plots of Appendix E portray actual RFETS groundwater data, the control limits "h" and "SCL" indicated on the plots have no significance to RFETS.

4.4.2 Normality Testing of RFETS Data

The groundwater quality data were tested for normality using the Shapiro-Wilk test for sample sizes <50, and the Shapiro-Francia test for larger data sets. A total of 48 normality tests were conducted using WQStat Plus. Output from these normality tests may be viewed in Appendix E. Table 4-1 summarizes the test results for data classified by analyte, well, and length of season. Twenty eight of the tests found insufficient evidence to reject the null hypothesis of normality. The remaining 20 tests found statistically significant evidence of non-normality at 95% confidence. Although these data are not a comprehensive evaluation of the normality of groundwater data at RFETS, the preliminary evidence is that roughly one third of RFETS groundwater quality data are not normally distributed. Therefore parametric tests should not be applied to this groundwater quality data without verifying the statistical assumptions supporting the test. In the absence of such verification, nonparametric methods should be used for trend analysis.

4.4.3 Seasonality Testing of RFETS Data

Selection of a trend analysis procedure is partly based on whether or not seasonality influences the data. For example, in the absence of seasonality the Mann-Kendall test for trend may be used on groundwater data. However, if seasonal effects are significant on water quality, then the Seasonal-Kendall test should be used. Alternatively, in the presence of seasonality, the Mann-Kendall test can be used on deseasonalized data.

The groundwater quality data from RFETS were tested for seasonality using the nonparametric Kruskal-Wallis test. The null hypothesis of this test is that each season has the same median concentration of a given analyte. Testing for seasonality via the Kruskal-Wallis test requires a minimum sample size of four data points per season (i.e., >= four years of seasonal data). Twenty eight sample sets contained sufficient data to be tested for seasonality. Statistically significant evidence of seasonality in RFETS groundwater was found at 95% confidence for the following three sets of data (Table 4-2).

- Well B206989 nitrate/nitrite data based on quarterly sampling seasons (B206989 is a RCRA well sampled quarterly; however it is frequently dry.)
- Well B206989 nitrate/nitrite data based on semiannual sampling seasons
- Well 06091 uranium-234 activity data based on quarterly seasons.

The above evidence of seasonality in groundwater at RFETS serves only to support the use of the Seasonal-Kendall test, or the use of the Mann-Kendall test on deseasonalized data. Data tested for seasonality represent a small percentage of the more than one million records of groundwater quality data collected over 18 years at RFETS. Therefore, on the basis of the above seasonality testing no conclusions should be reached regarding the nature and extent of seasonality in UHSU groundwater at RFETS.

58

Table 4-2 indicates that only 1 out of 15 seasonality tests was statistically significant (nitrate/nitrite at well B206989) for the semiannual data. Two out of 13 tests were statistically significant for the quarterly data. However, because of missing data and small sample sets, no conclusions should be drawn regarding the detectability of seasonality in semiannual sampling versus quarterly sampling. There has been limited monthly groundwater quality sampling at RFETS, but there were insufficient monthly data in the test data to evaluate seasonality using monthly seasons.

4.4.4 Autocorrelation Testing of RFETS Data

Published literature regarding trend analysis frequently mentions that serial correlation can impact the power of a statistical test for trend. Therefore, the RFETS groundwater data were tested for serial correlation by using the Rank Von Neumann test. Tests for serial correlation are also effected by the presence of seasonality and/or trend (IDT, 1998). Therefore, WQStat Plus was used to first deseasonalize the data and then detrend the data using Sen's slope trend estimate. The null hypothesis (H_0) is that no serial correlation is present (i.e., the data are independent). There is statistically significant evidence to reject H_0 when the computed statistic $R_v <$ the tabulated critical value at a given significance level.

The results of 30 Rank Von Neumann tests are found in Appendix E, and these results are tabulated on Table 4-3. This table indicates that 10 of the 30 tests had statistically significant evidence of serial correlation at 95% confidence. Monthly sampling should have higher autocorrelation than quarterly sampling, and quarterly sampling should have higher autocorrelation than semiannual sampling. However, this is not observed in the RFETS data, probably because of missing quarterly data, and even larger percentages of missing monthly observations. Examining the test results by season, Table 4-3 contains the following results,

- Semiannual seasons – four tests indicated autocorrelation, while 10 tests did not
- Quarterly seasons – five tests indicated autocorrelation, while 10 tests did not
- Monthly seasons – one test indicated autocorrelation.

Only one monthly test was performed because of insufficient data. The evidence of serial correlation in the semiannual data is only slightly less frequent than it is in the quarterly data. Semiannual groundwater sampling is proposed for post-closure monitoring at RFETS.

The Seasonal-Kendall test is robust against seasonality, non-normality, and can be used with some nondetects (censoring), or missing values (Hirsch et al., 1982). However, the Seasonal-Kendall is not robust against serial dependence (Hirsch and Slack, 1984). A modified version of the Seasonal-Kendall test was developed to be robust against serial dependence (Hirsch and Slack, 1984). The modified test does not work well with periods of record of less than 10 years, and it is less powerful than the original Seasonal-Kendall when the data are in fact independent (no serial correlation; Hirsch and Slack, 1984).

The modified Seasonal-Kendall test is more complex, and at the time of writing, software has not been located to evaluate it on RFETS data. Because the modified Seasonal-Kendall test was not evaluated and requires at least 10 years of data, it not recommended for post-closure monitoring.

60

Table 4-1 Normality Test Results for RFETS Groundwater Data

Seasons Per Year	Analyte	Well	Sample Size (N)	Calculated Shapiro-Wilk (Shapiro-Francia if N > 49)	Critical Value	Normal at 95% Confidence Y=Yes N=No
2	Chloroform	23296	15	0.883	0.881	Y
2	CT	06091	24	0.959	0.916	Y
2	Nitrate/Nitrite	70193	19	0.917	0.901	Y
2	Nitrate/Nitrite	06091	25	0.94	0.918	Y
2	Nitrate/Nitrite	B206989	22	0.959	0.911	Y
2	PCE	02291	19	0.915	0.901	Y
2	PCE	P114889	9	0.855	0.901	Y
2	TCE	00491	23	0.901	0.914	N
2	TCE	3586	32	0.759	0.93	N
2	TCE	43392	21	0.774	0.908	N
2	U-234	10194	18	0.916	0.897	Y
2	U-234	06091	25	0.913	0.918	N
2	U-234	B210489	16	0.934	0.887	Y
2	U-234	P415889	5	0.968	0.762	Y
2	Vinyl Chloride	3586	30	0.795	0.927	N
2	Vinyl Chloride	P115689	10	0.928	0.842	Y
4	Chloroform	23296	19	0.891	0.901	N
4	CT	06091	35	0.957	0.934	Y
4	Nitrate/Nitrite	70193	33	0.262	0.931	N
4	Nitrate/Nitrite	06091	35	0.948	0.934	Y
4	Nitrate/Nitrite	B206989	26	0.948	0.92	Y
4	PCE	02291	26	0.896	0.92	N
4	PCE	P114889	14	0.913	0.874	Y
4	TCE	00491	30	0.894	0.927	N
4	TCE	3586	52	0.734	0.955	N
4	TCE	43392	26	0.781	0.92	N
4	U-234	10194	21	0.944	0.908	Y
4	U-234	06091	33	0.943	0.931	Y
4	U-234	B210489	30	0.96	0.927	Y
4	U-234	P415889	8	0.669	0.818	N
4	Vinyl Chloride	3586	48	0.821	0.947	N
4	Vinyl Chloride	P115689	15	0.917	0.881	Y
12	Chloroform	23296	23	0.891	0.914	N
12	CT	06091	38	0.962	0.938	Y
12	Nitrate/Nitrite	70193	36	0.254	0.935	N
12	Nitrate/Nitrite	06091	38	0.942	0.938	Y
12	Nitrate/Nitrite	B206989	28	0.958	0.924	Y
12	PCE	02291	25	0.94	0.918	Y
12	PCE	P114889	14	0.92	0.874	Y

Seasons Per Year	Analyte	Well	Sample Size (N)	Calculated Shapiro-Wilk (Shapiro-Francia if N > 49)	Critical Value	Normal at 95% Confidence Y=Yes N=No
12	TCE	00491	32	0.914	0.93	N
12	TCE	3586	53	0.729	0.957	N
12	TCE	43392	26	0.781	0.92	N
12	U-234	10194	21	0.874	0.908	N
12	U-234	06091	35	0.953	0.934	Y
12	U-234	B210489	29	0.958	0.926	Y
12	U-234	P415889	8	0.82	0.818	Y
12	Vinyl Chloride	3586	49	0.844	0.947	N
12	Vinyl Chloride	P115689	15	0.917	0.881	Y

62

Table 4-2 Seasonality Test Results for RFETS Groundwater Data

Seasons Per Year	Analyte	Well	Kruskal-Wallis H statistic (adjusted for ties)	Tabulated Chi-Squared Value	Seasonality at 95% Confidence Y = Yes N = No
2	PCE	02291	0.482	3.841	N
2	PCE	P114889	0.060	3.841	N
2	CT	06091	1.342	3.841	N
2	U-234	06091	0.539	3.841	N
2	U-234	10194	0.955	3.841	N
2	U-234	B210489	0.070	3.841	N
2	Nitrate/Nitrite	06091	0.074	3.841	N
2	Nitrate/Nitrite	70193	0.007	3.841	N
2	**Nitrate/Nitrite**	**B206989**	**9.129**	**3.841**	**Y**
2	Chloroform	23296	0.483	3.841	N
2	Vinyl Chloride	3586	0.000	3.841	N
2	Vinyl Chloride	P115689	0.011	3.841	N
2	TCE	00491	0.095	3.841	N
2	TCE	3586	0.174	3.841	N
2	TCE	43392	0.000	3.841	N
4	CT	06091	1.928	7.815	N
4	**U-234**	**06091**	**11.334**	**7.815**	**Y**
4	U-234	B210489	3.930	7.815	N
4	Nitrate/Nitrite	06091	0.578	7.815	N
4	Nitrate/Nitrite	70193	4.235	7.815	N
4	**Nitrate/Nitrite**	**B206989**	**10.554**	**7.815**	**Y**
4	Vinyl Chloride	3586	**6.088**	**7.815**	N
4	Vinyl Chloride	P115689	0.879	7.815	N
4	PCE	02291	2.125	7.815	N
4	TCE	00491	1.850	7.815	N
4	TCE	3586	1.700	7.815	N
4	TCE	43392	**6.526**	**7.815**	N
4	Chloroform	23296	0.885	7.815	N

63

Table 4-3 Serial Correlation Test Results for RFETS Groundwater Data

Seasons Per Year	**Analyte**	**Well**	**Rank Von Neumann Statistic Rv**	**Table Critical Value**	**Significant at 95% Confidence Y = Yes N = No**
2	Chloroform	23296	2.24	1.19	N
2	CT	06091	1.27	1.35	Y
2	Nitrate/Nitrite	06091	1.31	1.36	Y
2	Nitrate/Nitrite	70193	1.54	1.27	N
2	Nitrate/Nitrite	B206989	1.89	1.32	N
2	PCE	02291	2.24	1.27	N
2	TCE	00491	1.95	1.33	N
2	TCE	3586	2.61	1.43	N
2	TCE	43392	1.16	1.31	Y
2	U-234	06091	3.38	1.36	N
2	U-234	10194	1.94	1.26	N
2	U-234	B210489	1.39	1.21	N
2	Vinyl Chloride	3586	0.97	1.41	Y
2	Vinyl Chloride	P115689	1.65	1.04	N
4	Chloroform	23296	2.60	1.27	N
4	CT	06091	0.82	1.455	Y
4	Nitrate/Nitrite	06091	1.50	1.455	N
4	Nitrate/Nitrite	70193	2.24	1.44	N
4	Nitrate/Nitrite	B206989	1.63	1.37	N
4	PCE	02291	2.41	1.37	N
4	PCE	P114889	1.84	1.17	N
4	TCE	00491	1.78	1.41	N
4	TCE	3586	1.48	1.548	Y
4	TCE	43392	0.75	1.37	Y
4	U-234	06091	1.85	1.44	N
4	U-234	10194	2.21	1.31	N
4	U-234	B210489	1.38	1.41	Y
4	Vinyl Chloride	3586	1.05	1.53	Y
4	Vinyl Chloride	P115689	2.00	1.19	N
12	CT	06091	1.14	1.48	Y

64

4.5 Trend Analysis Results from WQStat Plus

Test data sets containing groundwater quality data from RFETS were processed by program WQStat Plus. The objective was to compare three nonparametric trend analysis procedures, and associated slope estimation methods. The three procedures are:

- Mann-Kendall test for trend applied to unadjusted groundwater concentration data;
- Mann-Kendall test for trend applied to previously deseasonalized concentration data; and
- Seasonal-Kendall test for trend applied to unadjusted groundwater concentration data.

Deseasonalization was done following the method of EPA (1989). Each deseasonalized concentration equals its unadjusted concentration plus the grand mean, minus the seasonal mean. Deseasonalization is also discussed by IDT (1998).

The output from the above tests is contained in Appendix E. Output from procedure 3, the Seasonal-Kendall test, is entitled "Seasonal Kendall Slope Estimator" in Appendix E. The well name is printed under the title block, and the computed Seasonal-Kendall statistic is labeled "Z." Negative values of Z indicate a negative slope or downward trend, if the test is statistically significant at some confidence level (95% confidence is used in this report). Positive values of Z indicate a positive slope or upward trend, if the test is statistically significant.

Results from Procedure 1 (the Mann-Kendall on unadjusted data) are entitled "Sen's Slope Estimator" in Appendix E. The test is significant when the calculated Mann-Kendall statistic is greater than the critical value at 95% confidence (shown on the output as an alpha value of 0.05). Appendix E results from Procedure 2 (the Mann-Kendall on deseasonalized data) are entitled "Sen's Slope Estimator (Alt. Values)."

The test results (at 95% confidence) of the above three procedures have been tabulated in Table 4-4. The left-hand column shows the number of seasons assumed in creation of the data set. The results of Procedure 1 are tabulated 7 columns from the left as "U" for upward trend, "D" for downward trend, or "N" for no trend. These trend results are given at the 95% confidence level. Column 8 data is the steepness of the trend based on Sen's nonparametric slope estimator. This estimator has units of ug/L per year for non-radionuclides, and pCi/L per year for U-234.

Table 4-4 column 11 (from left) holds the results of procedure 2, the Mann-Kendall test on deseasonalized data. Again the Sen slope estimator follows it in column 12. Empty cells in the table indicate that there were insufficient data points to deseasonalize the data. At least four points are required per season.

65

Fifteen columns from left in Table 4-4 are the Seasonal-Kendall test results. Column 16 holds the related slope estimator results. More subjective, visual estimates of trend based on inspection of seasonality plots, are listed in the fifth column from the right side of the table.

The four right-most columns of Table 4-4 compare the trends predicted by the various methods as a crude indicator of relative performance. Specifically, the fourth column from the right compares the results of the ordinary Mann-Kendall test with its counterpart on deseasonalized data. This indicates that out of 33 comparisons, 30 of the tests (91%) gave the same predictions. Predictably the tests differed for quarterly seasons of nitrate/nitrite at Well B206989, and U-234 at Well 06091, which were previously shown to have statistically significant seasonality.

Column three (from right) compares 24 Seasonal-Kendall tests with their Mann-Kendall counterparts (the latter using deseasonalized data). Twenty one of the 24 tests (88%) gave the same results. Given this similar performance, it is simpler to use the Seasonal-Kendall for post-closure monitoring, and avoid the need to deseasonalize the concentration data.

Steep, highly linear data trends are easy to identify. Human interpretation of trends becomes more subjective if the smooth is highly nonlinear, or as the trend slope approaches zero. However, it is instructive to compare trend detection by humans (the visual trend result) with the Seasonal-Kendall test result at 95% confidence. This comparison is done in column 2 (from right). Again, 21 of 24 tests (88%) agree on the trend result.

Column 1 from right is a similar comparison of the visual trends against the Mann-Kendall test with deseasonalized data. Only 82% of these M-K tests agreed with the human interpretation. This is a second reason to prefer the Seasonal-Kendall for post-closure monitoring and to dispense with the need to deseasonalize.

In conclusion, all three nonparametric trend analysis methods worked quite well with the RFETS groundwater data. All three statistical methods were usually in agreement with the more subjective trend identifications made by humans. The Seasonal-Kendall test agreed very closely (88%) with the Mann-Kendall test when the latter test used deseasonalized data. Given this similar performance, it is simpler to use the Seasonal-Kendall for post-closure monitoring, and avoid the need to deseasonalize the concentration data.

66

Table 4-4 Comparison of Seasonal-Kendall and Mann-Kendall Tests.

Number of Seasons Per Year	Analyte	Well	Sample Size (N)	Ordinary Mann-Kendall (M-K) Test Statistic	Ordinary M-K Critical Value (Table Value)	Ordinary M-K Test Result at 95% Confidence[a]	Ordinary M-K Sen's Slope Estimator (ug/L per year or pCi/L per year)	Deseasonalized Mann-Kendall Test Statistic	Deseasonalized M-K Critical Value (Table Value)	Deseasonalized M-K test Result at 95% Confidence[a]	Deseasonalized M-K Sen's Slope Estimator (ug/L per year or pCi/L per year)	Seasonal Kendall Statistic Z	S-K Critical Value (Table Value)	S-K Test Result at 95% Confidence[a]	S-K Slope Estimator (ug/L per year or pCi/L per year)	Visual Estimate of Trend from Seasonality Plot[a]	Compare Ordinary M-K & Deseasonalized M-K Result[b]	Compare S-K with Deseasonalized M-K Result[b]	Compare S-K Result with Visual Trend Result[b]	Compare Deseasonalized M-K Result with Visual Trend Result[b]
2	CF	23296	15	6	41	N	0.238	12	41	N	0.829	0.48	1.96	N	0.829	N	A	A	A	A
2	CT	06091	24	110	81	U	0.37	112	81	U	0.366	2.795	1.96	U	0.37	U	A	A	A	A
2	Nitrate/ Nitrite	70193	19	96	58	U	118.7	98	58	U	131.2	2.939	1.96	U	119.2	U	A	A	A	A
2	Nitrate/ Nitrite	06091	25	75	85	N	178.6	77	85	N	183.3	1.46	1.96	N	153.3	N	A	A	A	A
2	Nitrate/ Nitrite	B206989	22	-12	-71	N	-271.9	-11	-71	N	-260.5	-0.275	1.96	N	-266.2	N	A	A	A	A
2	PCE	02291	19	98	58	U	483.6	97	58	U	468.3	2.987	1.96	U	498.2	U	A	A	A	A
2	PCE	P114889	9	20	20	U	16.7	20	20	U	17.7	1.788	1.96	N	16.1	N	A	D	A	D
2	TCE	00491	23	-117	-76	D	-3.46	-117	-76	D	-3.519	-2.992	1.96	D	-3.41	D	A	A	A	A
2	TCE	3586	32	-295	-123	D	-0.289	-275	-123	D	-0.346	-4.636	1.96	D	-0.294	D	A	A	A	A
2	TCE	43392	21	-127	-66	D	-0.061	-127	-66	D	-0.06	-3.817	1.96	D	-0.057	D	A	A	A	A
2	U-234	10194	18	-69	-53	D	-0.101	-69	-53	D	-0.101	-2.464	1.96	D	-0.113	N	A	A	D	D
2	U-234	06091	25	57	85	N	0.044	56	85	N	0.056	1.413	1.96	N	0.044	N	A	A	A	A
2	U-234	B210489	16	12	45	N	0.209	12	45	N	0.208	0.514	1.96	N	0.226	N	A	A	A	A
2	U-234	P415889	5	-2	-10	N	-0.179	-4	-10	N	-0.126	-----	-----	-----	-----	N	A	-----	-----	A
2	VC	3586	30	-336	-112	D	-33	-312	-112	D	-33	-5.822	1.96	D	-33.3	D	A	A	A	A
2	VC	P115689	10	35	23	U	13	35	23	U	12.5	2.598	1.96	U	13.1	U	A	A	A	A
4	CF	23296	19	-20	-58	N	-0.989	-3	-58	N	-0.06	-0.253	1.96	N	-1.35	N	A	A	A	A
4	CT	06091	35	292	140	U	0.483	299	140	U	0.458	3.731	1.96	U	0.419	U	A	A	A	A

Review Exemption: CEX-105-01

Number of Seasons Per Year	Analyte	Well	Sample Size (N)	Ordinary Mann-Kendall (M-K) Test Statistic	Ordinary M-K Critical Value (Table Value)	Ordinary M-K Test Result at 95% Confidence[a]	Ordinary M-K Sen's Slope Estimator (ug/L per year or pCi/L per year)	Deseasonalized Mann-Kendall Test Statistic	Deseasonalized M-K Critical Value (Table Value)	Deseasonalized M-K test Result at 95% Confidence[a]	Deseasonalized M-K Sen's Slope Estimator (ug/L per year or pCi/L per year)	Seasonal Kendall Statistic Z	S-K Critical Value (Table Value)	S-K Test Result at 95% Confidence[a]	S-K Slope Estimator (ug/L per year or pCi/L per year)	Visual Estimate of Trend from Seasonality Plot[a]	Compare Ordinary M-K & Deseasonalized M-K Result[b]	Compare S-K with Deseasonalized M-K Result[b]	Compare S-K Result with Visual Trend Result[b]	Compare Deseasonalized M-K Result with Visual Trend Result[b]
4	Nitrate/ Nitrite	70193	33	246	128	U	115.7	90	128	U	128.3	3.62	1.96	U	104.8	U	A	A	D	D
4	Nitrate/ Nitrite	06091	35	128	140	N	154.6	156	140	U	181.8	1.492	1.96	N	153.3	N	D	D	A	D
4	Nitrate/ Nitrite	B206989	26	-29	-90	N	-390.9	-26	-90	N	-279.5	-----	-----	-----	-----	N	A	-----	-----	A
4	PCE	02291	26	144	90	U	448.9	123	90	U	344.6	-----	-----	-----	-----	U	A	-----	-----	A
4	PCE	P114889	14	38	37	U	19.6	[illegible]9	37	U	11.9	-----	-----	-----	-----	N	A	-----	-----	D
4	TCE	00491	30	-271	-112	D	-6.36	-270	-112	D	-6.08	-5.144	1.96	D	-6.79	D	A	A	A	A
4	TCE	3586	52	-6.48[c]	-1.96	D	-0.495	-4.93[c]	-1.96	D	-0.521	-5.756	1.96	D	-0.454	D	A	A	A	A
4	TCE	43392	26	-192	-90	D	-0.06	-120	-90	D	-0.053	-----	-----	-----	-----	N	A	-----	-----	D
4	U-234	10194	21	-108	-66	D	-0.097	-[illegible]8	-66	N	-0.028	-----	-----	-----	-----	N	D	-----	-----	A
4	U-234	06091	33	133	128	U	0.052	86	128	N	0.041	2.173	1.96	U	0.061	N	D	D	D	A
4	U-234	B210489	30	-43	-112	N	-0.343	-[illegible]7	-112	N	-0.449	-0.534	1.96	N	-0.449	N	A	A	A	A
4	U-234	P415889	8	-6	-17	N	-0.144	[illegible]	17	N	-0.007	-----	-----	-----	-----	N	A			A
4	VC	3586	48	-6.94[c]	-1.96	D	-38.6	-5.[illegible]2[c]	-1.96	D	-35.5	-6.209	1.96	D	-36.2	D	A	A	A	A
4	VC	P115689	15	73	41	U	14.6	6[illegible]	41	U	15	-----	-----	-----	-----	U	A	-----	-----	A
12	CF	23296	23	-10	-76	N	-0.119	-----	-----	-----	-----	-----	-----	-----	-----	N	-----	-----	-----	
12	CT	06091	38	347	158	U	0.53	263	158	U	0.38	-----	-----	-----	-----	U	A	-----	-----	A
12	Nitrate/ Nitrite	70193	36	291	145	U	104.6	-----	-----	-----	-----	-----	-----	-----	-----	U	-----	-----	-----	-----
12	Nitrate/ Nitrite	06091	38	131	158	N	135.1	-----	-----	-----	-----	-----	-----	-----	-----	U	-----	-----	-----	-----
12	Nitrate/ Nitrite	B206989	28	-18	-101	N	-245.1	-----	-----	-----	-----	-----	-----	-----	-----	N	-----	-----	-----	-----

Number of Seasons Per Year	Analyte	Well	Sample Size (N)	Ordinary Mann-Kendall (M-K) Test Statistic	Ordinary M-K Critical Value (Table Value)	Ordinary M-K Test Result at 95% Confidence[a]	Ordinary M-K Sen's Slope Estimator (ug/L per year or pCi/L per year)	Deseasonalized Mann-Kendall Test Statistic	Deseasonalized M-K Critical Value (Table Value)	Deseasonalized M-K test Result at 95% Confidence[a]	Deseasonalized M-K Sen's Slope Estimator (ug/L per year or pCi/L per year)	Seasonal Kendall Statistic Z	S-K Critical Value (Table Value)	S-K Test Result at 95% Confidence[a]	S-K Slope Estimator (ug/L per year or pCi/L per year)	Visual Estimate of Trend from Seasonality Plot[a]	Compare Ordinary M-K & Deseasonalized M-K Result[b]	Compare S-K with Deseasonalized M-K Result[b]	Compare S-K Result with Visual Trend Result[b]	Compare Deseasonalized M-K Result with Visual Trend Result[b]
12	PCE	02291	25	147	85	U	492.2	-----	-----	-----	-----	-----	-----	-----	-----	U	-----	-----	-----	-----
12	PCE	P114889	14	44	37	U	17.1	-----	-----	-----	-----	-----	-----	-----	-----	N	-----	-----	-----	-----
12	TCE	00491	32	-262	-123	D	-5.22	-----	-----	-----	-----	-----	-----	-----	-----	D	-----	-----	-----	-----
12	TCE	3586	53	-6.60[c]	-1.96	D	-0.457	-----	-----	-----	-----	-----	-----	-----	-----	D	-----	-----	-----	-----
12	TCE	43392	26	-192	-90	D	-0.06	-----	-----	-----	-----	-----	-----	-----	-----	D	-----	-----	-----	-----
12	U-234	10194	21	-92	-66	D	-0.089	-----	-----	-----	-----	-----	-----	-----	-----	D	-----	-----	-----	-----
12	U-234	06091	35	152	140	U	0.057	-----	-----	-----	-----	-----	-----	-----	-----	N	-----	-----	-----	-----
12	U-234	B210489	29	-44	-106	N	-0.321	-----	-----	-----	-----	-----	-----	-----	-----	N	-----	-----	-----	-----
12	U-234	P415889	8	-3	-17	N	-0.149	-----	-----	-----	-----	-----	-----	-----	-----	N	-----	-----	-----	-----
12	VC	3586	49	-6.99[c]	-1.96	D	-38.3	-----	-----	-----	-----	-----	-----	-----	-----	D	-----	-----	-----	-----
12	VC	P115689	15	73	41	U	14.6	-----	-----	-----	-----	-----	-----	-----	-----	U	-----	-----	-----	-----

Notes:

[a] N = no significant trend; D = statistically significant downward trend; U = statistically significant upward trend.

[b] A = agrees; D = disagrees.

[c] Mann-Kendall critical value based on normal approximation method.

This page intentionally left blank.

4.6 Trend Analysis Results from Program TREND3

Previous Section 4.5 discussed trends visually identified from inspection of seasonality plots and compared to hypothesis tests of trend performed by WQStat Plus. However, there are other ways to evaluate trends. In the present Section, trends have been visually estimated from many of the LOWESS smooths, which were discussed in Section 4.3. These visual trends were compared to the Seasonal-Kendall test of trend performed by program TREND3 (Gilbert, 1987, Appendix B).

Another difference is that the nondetect results were used at the reported value in the WQStat Plus runs. Nondetects were included at one-half it's reported value in the TREND3 runs. The output of TREND3 is contained in Appendix F and the test results are tabulated on Table 4-5.

Table 4-5, column four (from left) shows that between 8 and 15 years of data were potentially available for trend testing. However, the 8th column indicates that substantial percentages of data were missing. Missing data reached a maximum of 89% for vinyl chloride at Well P115689, assuming monthly seasons. Inspection of Table 4-5 indicates that monthly seasons have the highest percentages of missing data, and semiannual seasons have the lowest. This is because relatively little monthly groundwater sampling is performed at RFETS. RCRA wells B206989 and 70193 are sampled quarterly, but show missing data because of limited availability of groundwater near the Present Landfill.

Table 4-5, column three (from left) lists the subjective visual estimate of trend for each analyte-well combination. Column three also notes data problems, such as outliers, observed in each data set. The 5th column indicates that semiannual, quarterly and monthly seasons were evaluated, and the 6th column shows the number of seasonal data points available. When the data sets were created a maximum of one data point was selected to represent each season.

The null hypothesis of the Seasonal-Kendall test is that there is no trend. When the absolute value of the computed Z statistic is greater than the critical value of 1.645, then there is statistically significant evidence of a trend at 95% confidence. If the Z score is positive the trend is upwards; if negative the trend is downwards. The 3rd column from the right on Table 4-5 indicates if a significant trend was detected by the Seasonal-Kendall test.

The right-hand column of Table 4-5 compares the visual trend prediction based on inspection of LOWESS smooths with the Seasonal-Kendall test results. Out of 21 comparisons, 18 agree (86%). This result is consistent with the 88% agreement found in Section 4.5 which compared Seasonal-Kendall tests run by WQStat Plus to visual trend identification based on seasonality plots. Much of the 14% disagreement between the Seasonal-Kendall and visual predictions based on smooths is probably the result of the subjectivity of trend identification by humans.

Table 4-5 indicates that the Seasonal-Kendall test generally predicts the same trend (or lack of trend), regardless of whether the data consist of two semiannual seasons, four quarterly seasons, or 12 monthly seasons. This was true despite the fact that much of the monthly seasonal data was missing from the data sets. This absence occurred because groundwater monitoring at RFETS is more commonly semiannual or quarterly.

71

72

Table 4-5 Comparison of Visual Trends Identified from LOWESS Smooths with Seasonal-Kendall Trend Results Computed by Program TREND3.

Analyte	Well	Visual Trend from LOWESS or Time Series Plot?	Maximum Years of Data Record	Assumed Seasons/ Year for Seasonal- Kendall (S-K) test	Number of Data Points Available for Testing	Approximate Number of Missing Seasonal Data Points	Percentage of Missing Seasonal Data	Seasonal-Kendall Computed Z	Trend Predicted by Seasonal-Kendall at 95% Confidence? (1 tailed test; critical Z= 1.645)	Type of Trend Predicted by S-K test	Did S-K Trend Identification agree with Visual?
nitrate/nitrite	B206989	No obvious trend	15	2	22	8	26.7	-0.275	No	None	Yes
nitrate/nitrite	B206989	No obvious trend	15	4	26	34	56.7	-0.283	No	None	Yes
nitrate/nitrite	B206989	No obvious trend	15	12	28	152	84.4	-0.125	No	None	Yes
nitrate/nitrite	70193	Slight upward trend in data; 1 extreme outlier	11	2	19	3	13.6	2.939	Yes	Upward Trend	Yes
nitrate/nitrite	70193	Slight upward trend in data; 1 extreme outlier	11	4	33	11	25.0	3.62	Yes	Upward Trend	Yes
nitrate/nitrite	70193	Slight upward trend in data; 1 extreme outlier	11	12	36	96	72.7	2.484	Yes	Upward Trend	Yes
vinyl chloride	P115689	Strong upward trend; relatively few data points; 5 year data gap.	11	2	10	12	54.5	2.598	Yes	Upward Trend	Yes
vinyl chloride	P115689	Strong upward trend; relatively few data points; 5 year data gap.	11	4	15	29	65.9	2.624	Yes	Upward Trend	Yes
vinyl chloride	P115689	Strong upward trend; relatively few data points; 5 year data gap.	11	12	15	117	88.6	1.508	No	None	No at 95% confidence; Yes at 93%
tetrachloroethene	02291	Moderate upward trend, but concentration data very noisy	13	2	19	7	26.9	2.987	Yes	Upward Trend	Yes
tetrachloroethene	02291	Moderate upward trend, but concentration data very noisy	13	4	26	26	50.0	2.379	Yes	Upward Trend	Yes

Review Exemption: CEX-105-01

13

Analyte	Well	Visual Trend from LOWESS or Time Series Plot?	Maximum Years of Data Record	Assumed Seasons/ Year for Seasonal- Kendall (S-K) test	Number of Data Points Available for Testing	Approximate Number of Missing Seasonal Data Points	Percentage of Missing Seasonal Data	Seasonal-Kendall Computed Z	Trend Predicted by Seasonal- Kendall at 95% Confidence? (1 tailed test; critical Z= 1.645)	Type of Trend Predicted by S-K test	Did S-K Trend Identification agree with Visual?
tetrachloroethene	02291	Moderate upward trend, but concentration data very noisy	13	12	25	131	84.0	2.111	Yes	Upward Trend	Yes
chloroform	23296	No obvious trend	8	2	15	1	6.3	0.480	No	None	Yes
chloroform	23296	No obvious trend	8	4	19	13	40.6	-0.253	No	None	Yes
chloroform	23296	No obvious trend	8	12	23	73	76.0	0.515	No	None	Yes
U-234	10194	Slight downward trend; low activity range in data; noisy data.	10	2	18	2	10.0	-2.464	Yes	Downward Trend	Yes
U-234	10194	Slight downward trend; low activity range in data; noisy data.	10	4	21	19	47.5	-1.345	No	None	No at 95% confidence; Yes at 91%
U-234	10194	Slight downward trend; low activity range in data; noisy data.	10	12	21	99	82.5	-0.112	No	None	No
carbon tetrachloride	06091	Upward trend strong for 8 years, then trend reversal & downwards out to 12 years of record.	12	2	24	0	0.0	2.795	Yes	Upward Trend	Yes
carbon tetrachloride	06091	Upward trend strong for 8 years, then trend reversal & downwards out to 12 years of record.	12	4	35	13	27.1	3.731	Yes	Upward Trend	Yes
carbon tetrachloride	06091	Upward trend strong for 8 years, then trend reversal & downwards out to 12 years of record.	12	12	38	106	73.6	1.935	Yes	Upward Trend	Yes

4.7 Trend Analysis Results Using An Exogenous Variable

Trend analysis of groundwater quality data normally examines analyte concentration as a function of elapsed time, measured from the dates of water sampling events. An external, independent variable such as net infiltration, or the variation in water table elevation, may influence the groundwater concentration data. This is postulated to happen from the mixing of two water quality types: infiltrating precipitation (relatively dilute concentrations due to short residence time), and groundwater with higher concentrations and longer residence time. It may be possible to mathematically model this influence and remove it from the data. Removal of this external variable may enhance our ability to detect a concentration versus time trend.

An exogenous variable is an independent variable whose value is determined externally to a statistical model. By contrast, an endogenous variable is a dependent variable whose value is determined within the model. As discussed earlier, a plausible exogenous variable for evaluation with groundwater quality data is the variation in water table elevation. It is postulated that the change in elevation reflects seasonal variations in groundwater recharge due to infiltrating precipitation, which may in turn affect analyte concentrations in UHSU groundwater at RFETS.

Historical water level measurements were retrieved from SWD for the wells and sampling events used in the trend testing quarterly data (Appendix C). Water levels are measured at RFETS as the depth to water (in decimal feet) below the top of the well casing. This depth to water usually varies seasonally, and between sampling events. The mean water level depth was computed for each well using all available water level measurements. The water level deviation for each well and sampling event was computed as the mean depth minus the water level measured prior to sampling. Therefore, positive water level changes indicate a water table elevated above the mean, while negative changes indicate a depressed water table.

Because each well has a different depth from the ground surface to its mean water table, and the wells were installed at various surface elevations, the physical elevation of the water table is not of particular interest here. It was convenient to work with changes in water table depth measured relative to the top of well casing.

For each combination of analyte and well, concentration data (measured quarterly), were plotted on the vertical axis (y-axis) versus the corresponding water level deviations on the horizontal axis (see scatterplot Figures 4-20, 4-21, 4-22). A second order LOWESS smooth was then fitted to each scatterplot by program Robust Fit. These smooths are the curves or line segments shown on the plots. The smooth is assumed to indicate the general effect on concentration of the changes in water level elevation.

Figures 4-21 and 4-22 indicate wide scatter in the concentration versus water level deviation data. The horizontal point band of Figure 4-20 indicates that nitrate/nitrite concentration is relatively insensitive to 10.4 feet of change in water levels at Well 70193. The coordinates of each fitted smooth may be exported from Robust Fit to a spreadsheet.

The effect of water table fluctuation on concentration is removed by computing residual concentrations. The residual concentration (delta C) is computed as each measured concentration minus the corresponding y-value of the smooth. These concentration residuals may then be plotted versus time, and analyzed for trend using the Seasonal-Kendall test in WQStat Plus.

Seasonality plots were generated in WQStat for TCE concentration residuals at wells 00491, and 3586 and for nitrate/nitrite at Well 70193. The Kruskal-Wallis test failed to find evidence of seasonality in any of these plots at 95% confidence.

The Seasonal-Kendall test was applied to the concentration residuals; the results are shown in Table 4-6. The absolute values of the Z's calculated on concentration residuals are smaller than the Z's calculated from concentration data which were not adjusted for water level changes. Therefore, the water level adjustments weakened the Seasonal-Kendall test's ability to predict concentration versus time trends. In fact, the TCE concentration residuals in Well 00491 failed to find evidence of a decreasing trend at 95% confidence. It did detect the decreasing trend at a lower confidence of 89%. These trend analysis comparisons were performed on a small sampling of analytes and wells so the evidence should be considered preliminary. In conclusion, preliminary evidence indicates that water level adjustment of analyte concentrations in RFETS groundwater fails to enhance detection of concentration versus time trends.

15

Review Exemption: CEX-105-01

Table 4-6 Comparison of Seasonal-Kendall Test Results using Either Unadjusted Concentration Data, or Adjusting it for Water Level Fluctuations.

Analyte	Well	Z Calculated on Adjusted Concentration Residuals	Z Calculated Previously on Unadjusted Concentration Data	S-K Trend on adjusted data At 95% confidence	S-K Trend on Unadjusted Data At 95% Confidence	Visual Trend on Unadjusted Data
TCE	00491	-1.613	-5.144	None	Down	Down
TCE	3586	-1.966	-5.756	Down	Down	Down
Nitrate/ nitrite	70193	3.144	3.62	Up	Up	Up

76

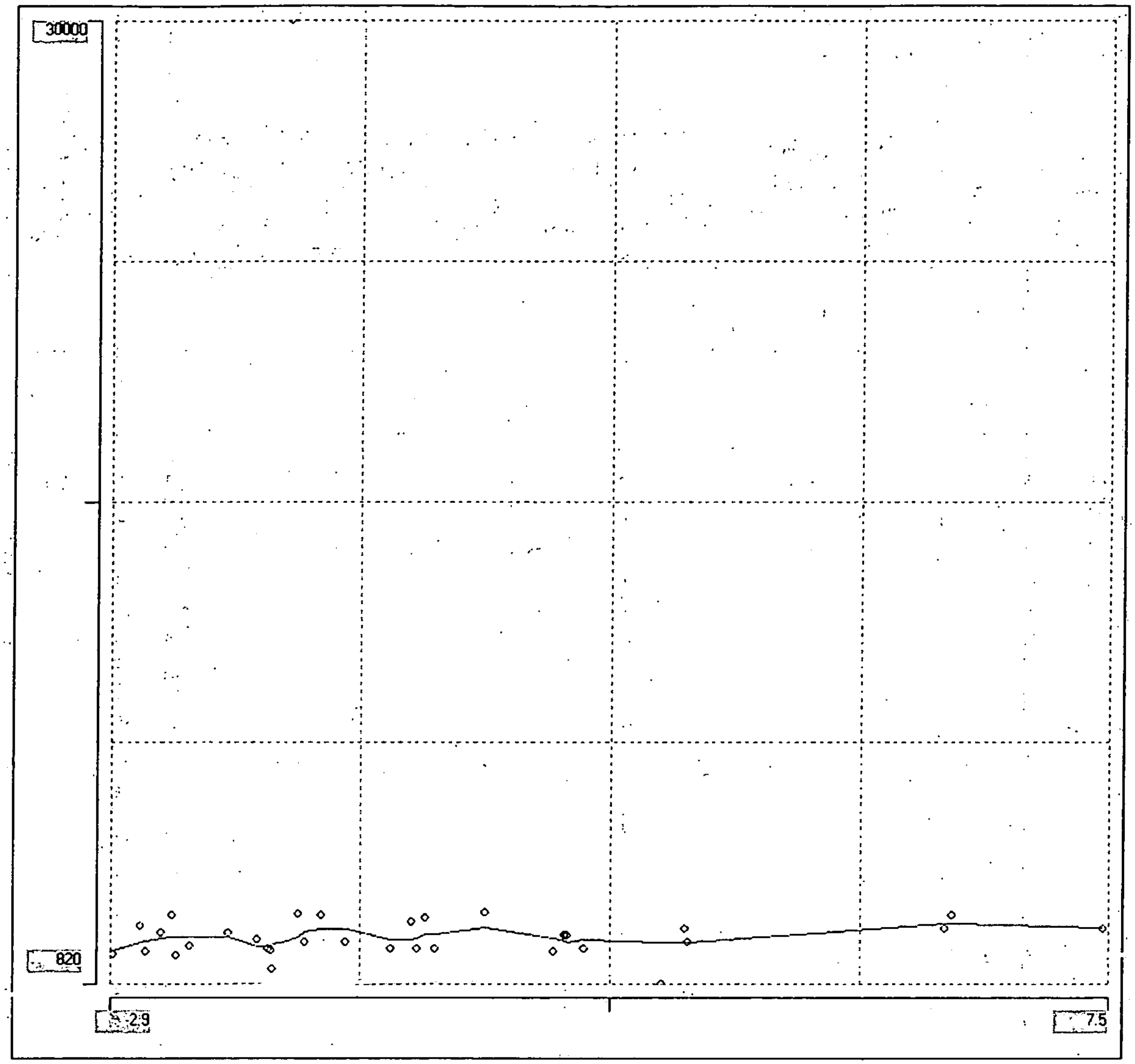

Figure 4-20 Smooth of Nitrate/Nitrite (as N) Concentration (y-axis) versus Water Level Deviation from the Mean (x-axis) at Well 70193.

The smooth ignores the 30 mg/L outlier. (2[nd] order smooth, n=33, half-window=6). The nitrate/nitrite concentrations do not appear to be strongly influenced by up to 10.4 feet of variation in groundwater level at this well.

77

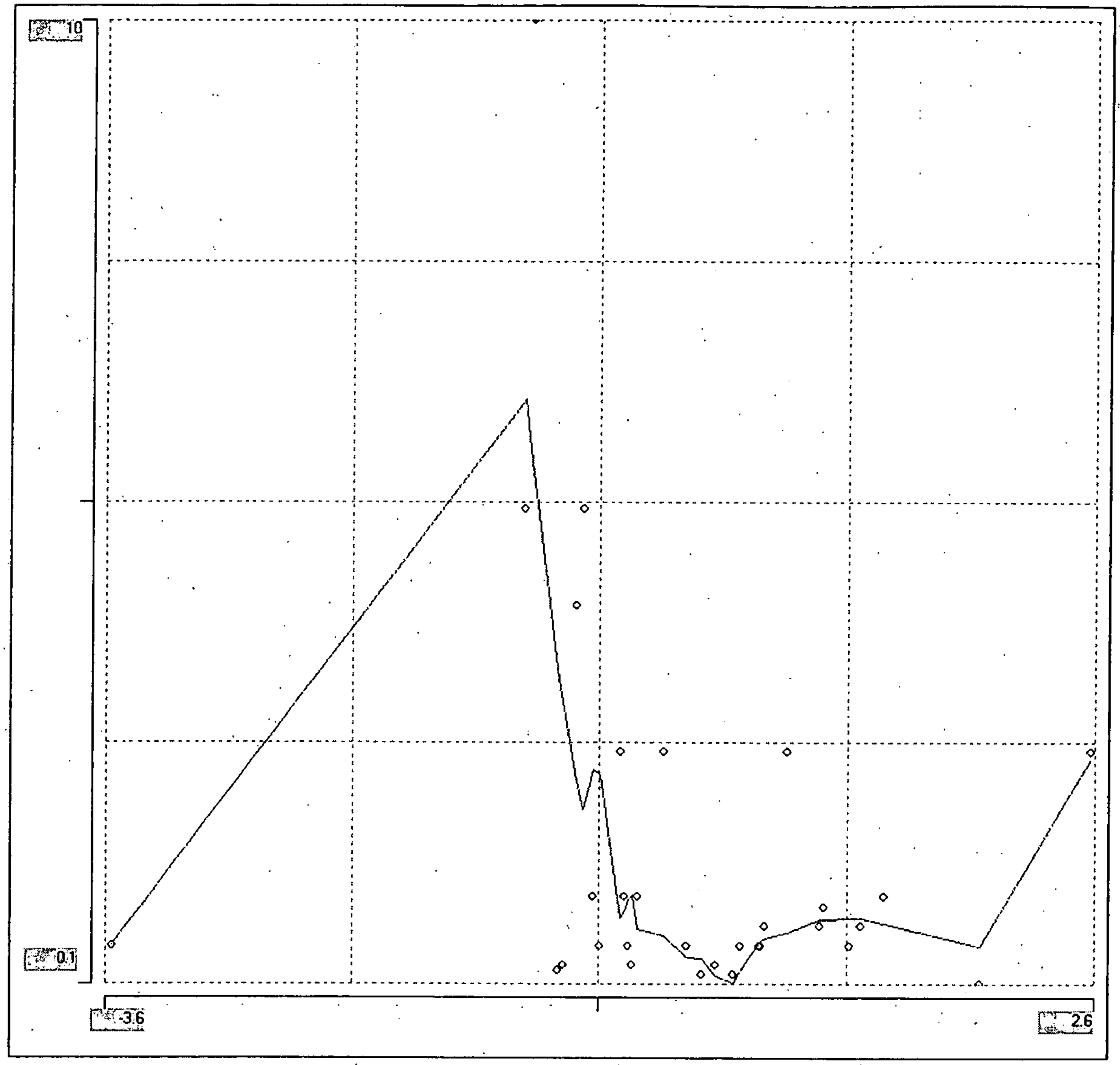

Figure 4-21 Smooth of TCE Concentration (y-axis) versus Water Level Deviation from the Mean (x-axis) at Well 3586.

This is a 2nd order smooth, n=32, half-window=6. The TCE concentration does not appear to be affected by up to 6.2 feet of water level change at this well.

78

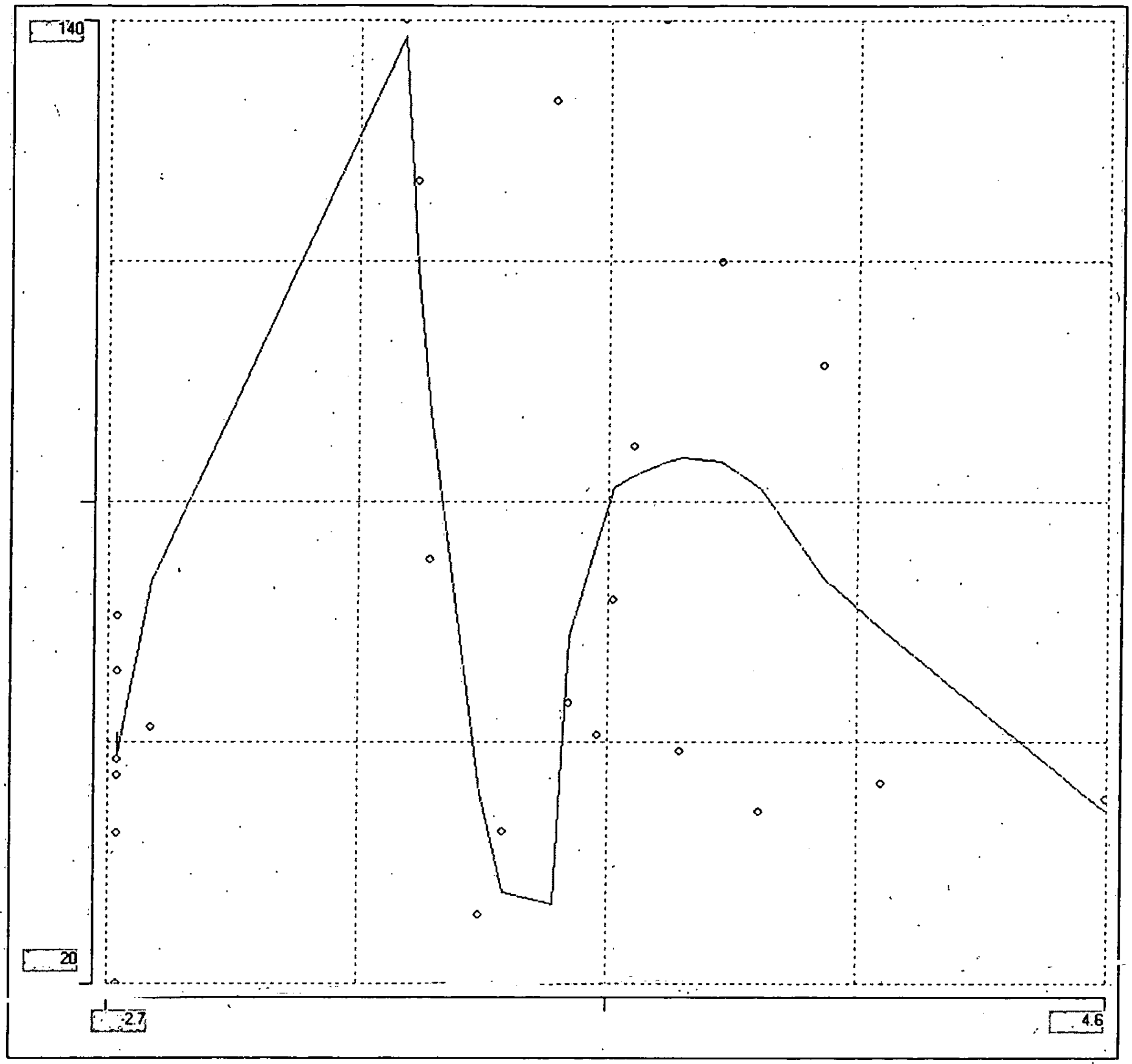

Figure 4-22 Smooth of TCE Concentration (y-axis) versus Water Level Deviation from the Mean (x-axis) at Well 00491.

This is a 2nd order smooth, n=26, half-window=6.

79

4.8 Application of Trending Methodology at RFETS

This report has been limited to a review of water quality trending methods. This work represents only a single aspect in the design of a post-closure monitoring program for RFETS. Numerous technical decisions remain to be made regarding data collection and interpretation. Examples include decisions about the list of analytes to be monitored, suitable analytical methods, required detection limits, selection of appropriate monitoring well locations, number of wells, screened interval, and sampling frequency.

A well located far downgradient of the IA is expected to contain groundwater with nondetect or very low concentrations of manmade COCs, and background levels of naturally occurring COCs. Groundwater samples collected from that well may yield VOC data that is 99% or 100% nondetects. Such data cannot be meaningfully trend tested. Trending analysis also requires some minimum number of data points collected at a regular sampling frequency, over a monitoring period of at least four years.

These technical decisions will be documented in a sampling and analysis plan for post-closure groundwater monitoring at RFETS, or a similar document, yet to be developed. The plan should contain all of the elements of a traditional sampling and analysis plan for environmental monitoring. It should also include a flowchart of the statistical decisions and logic to be followed in processing and testing the post-closure monitoring data.

This page intentionally left blank.

5 SUMMARY AND CONCLUSIONS

This investigation has reviewed much of the published literature regarding trend analysis methods. Emphasis was placed on trend analysis of environmental water quality data. The literature review indicates a number of data properties that may influence the selection of a statistical trend analysis method and its ability to correctly recognize a trend. These data properties include:

- The statistical distribution of the data (the population from which it came);
- Censored data and multiple detection limits in the data;
- Serial correlation in the data;
- Seasonality and climatic effects on water quality;
- Observations missing from the data;
- Multiple observations per combination of analyte-monitoring well-season;
- Temporally variable data, inconsistent sampling or analytical procedures;
- Data outliers; and
- The period of data record. Is there enough data, or is there too much data?

This investigation considered step trend and monotonic trend identification. Monotonic trend analysis is the most common type, and it is performed unless there is reason to anticipate a step trend in water quality. It is concluded that numerous D&D and remediation events at RFETS (e.g., groundwater treatment systems) may affect local downgradient groundwater quality, but that changes will likely be gradual.

Most of the published investigations of water quality trends since 1980 have utilized nonparametric statistical tests for monotonic trend. Parametric methods of trend analysis require that the sample data are independent and were drawn from a normally distributed population. Parametric methods that involve computation of the sample mean and standard deviation are more seriously impacted by data outliers than are nonparametric methods. Experience with groundwater quality data has also indicated that these data are frequently non-normal. This shifted the focus of the report to selecting some candidate nonparametric methods for further evaluation.

Three nonparametric trend analysis methods that are widely used in the water quality literature were selected as candidate methods for further testing using groundwater quality data collected at RFETS. The candidate methods were:

- Mann-Kendall test for trend on unadjusted concentration data. (Sen's slope estimator method was used with the Mann-Kendall to estimate trend magnitudes);

81

- Mann-Kendall test for trend on deseasonalized concentration data. (Deseasonalization was performed by the method of EPA (1989)); and

- Seasonal-Kendall test for trend on unadjusted concentration data. (The Seasonal-Kendall slope estimation method was used with this test.).

Groundwater data for testing these methods was drawn from the "Groundwater Superset." Test data contained examples of many of the data properties listed above, e.g., outliers and missing data. Data sets were created to compare trend test results for three groundwater sampling intervals (seasons): semiannual, quarterly, and monthly.

Shapiro-Wilk tests indicated that about 40% of the groundwater data were not normally distributed. This supported the decision to use nonparametric trend analysis methods.

The three candidate nonparametric methods gave similar trend predictions regardless of whether the input data were defined as semiannual seasons, quarterly seasons, or monthly seasons. This was not expected since much of the data were missing for monthly seasons. The methods all worked consistently with the RFETS groundwater test data despite the presence of data issues such as outliers.

The Seasonal-Kendall test is preferred over the ordinary Mann-Kendall test for post-closure monitoring at RFETS for the following reasons:

- Statistically significant Seasonality was identified at 95% confidence in some of the RFETS groundwater test data. The Mann-Kendall on unadjusted data does not account for this.

- The Seasonal-Kendall test agreed closely (88% of the time) with the Mann-Kendall test when the latter was used on deseasonalized concentration data. Thus we should use the Seasonal-Kendall and avoid the need to deseasonalize the groundwater data.

- The Seasonal-Kendall test (used at 95% confidence) agreed well (86 to 88%) with the more subjective trend identifications made by human inspection of seasonality plots, or of LOWESS smooths.

The Rank Von Neumann test was run to test for serial correlation in the groundwater data. Data were first deseasonalized and then detrended prior to running the Rank Von Neumann test, because it is sensitive to trend and seasonality. Statistically significant (at 95% confidence) serial correlation was found in some of the quarterly and semiannual data. The literature indicates that serial correlation is greatest at higher sampling frequencies. Therefore, semiannual groundwater sampling is preferred over quarterly or monthly sampling, for post-closure monitoring.

Measured deviations from mean water table elevation were used as an exogenous variable for trend analysis of groundwater. LOWESS smooths were used to adjust concentration data for groundwater for hypothetical impacts due to water table changes at RFETS. The Seasonal-Kendall test was applied to the

82

adjusted data (concentration residuals) and the results were compared with those based on unadjusted data. The absolute values of the Z statistics calculated on concentration residuals were smaller than the Z's calculated from concentration data which were not adjusted for water level changes. Therefore the water level adjustments weakened the ability of the Seasonal-Kendall test to predict concentration versus time trends. This preliminary evidence (based on a few well-analyte combinations) indicates that water level adjustment of analyte concentrations in RFETS groundwater fails to enhance detection of concentration versus time trends.

The literature says that nonparametric trend analysis methods, including the Seasonal-Kendall test, are not robust against serial correlation. However, despite the serial correlation found in the RFETS test data, the trend predictions of the Seasonal-Kendall test agreed very well with the visual trend predictions of humans. It is concluded that the Seasonal-Kendall test should work well on groundwater data collected semiannually for post-closure monitoring.

A possible alternative for post closure monitoring is to use a modified version of the Seasonal-Kendall test which compensates for serial correlation (Hirsch and Slack, 1984). However, this modified test requires at least 10 years of record, and is less powerful than the ordinary Seasonal-Kendall test when the data lack serial correlation. Therefore, if statistically significant serial correlation is not found to be abundant in the post-closure data, the ordinary unmodified Seasonal-Kendall remains the best monotonic trend test.

This page intentionally left blank.

6 REFERENCES

Alley, W.M., 1988, Using exogenous variables in testing for monotonic trends in hydrologic time series: Water Resources Research, V. 24, No. 11, pp. 1955-1961.

Antonopoulos, V.Z., Papamichail, D.M., and Mitsiou, K.A., 2001, Statistical and trend analysis of water quality and quantity data for the Strymon River in Greece, Hydrology and Earth System Sciences, v. 5, no. 4, pp. 679-691.

ASTM (American Society for Testing and Materials), 1975, Outlier Test Method Paper E178-75.

ASTM, 2002, E178-02 Standard Practice for Dealing with Outlying Observations.

Aziz, J.J., et al., 2003, Monitoring and Remediation Optimization System (MAROS) Software Version 2.0 User's Guide. Air Force Center for Environmental Excellence (AFCEE).

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

Cleveland, W.S., 1979, Robust locally weighted regression and smoothing scatterplots: J. Amer. Statist. Assn., v. 74, 829-836.

Conover, W.J., 1999, Practical Nonparametric Statistics, 3rd Ed. John Wiley & Sons. See the Sign Test on pp. 157-169, and the Cox and Stuart Test for Trend pp. 169-175.

Cox, D.R., and Stuart, A., 1955, Some quick sign tests for trend in location and dispersion: Biometrika, v. 42, no. 1/2, 80-95.

Daniels, H.E., 1950, Rank correlation and population models: J. Royal Statistical Soc., Series B, v. 12, no. 2, 171-191.

Dixon, W.J., 1953, Processing data for outliers: Biometrics, V. 9, p.74-89.

El-Shaarawi, A.H., Esterby, S.R., and Kuntz, K.W., 1983, A statistical evaluation of trends in the water quality of the Niagara River: J. Great Lakes Research, V. 9, p.234-240.

El-Shaarawi, A.H., and Niculescu, S.P., 1992, On Kendall's tau as a test of trend in time: Environmetrics, V. 3, p. 385-412.

EPA (U. S. Environmental Protection Agency), 1989, Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities. EPA/530-SW-89-026.

EPA, 1992, Methods for Evaluating the Attainment of Cleanup Standards. Volume 2 Groundwater. EPA230-R-92-014.

84

EPA, 2000, Guidance for Data Quality Assessment, Practical Methods for Data Analysis. EPA QA/G-9. QA00 Update. EPA/600/R-96/084. See Chapter 4.3 Tests for Trends.

EPA, CDPHE, and DOE, 1996, Final Rocky Flats Cleanup Agreement, July 19, 1996.

Fenelon, J.M. and Moreo, M.T., 2002, Trend Analysis of Groundwater Levels and Spring discharge in the Yucca Mountain Region, Nevada and California, 1960-2000. USGS WRI Rpt. 02-4178.

Gauthier, T.D., 2001, Detecting trends using Spearman's rank correlation coefficient: Environmental Forensics, V. 2, p. 359-362.

Gibbons, R.D., 1994, Statistical Methods for Groundwater Monitoring. John Wiley & Sons. 286 pp. See Chapter 9 Trend Analysis.

Gilbert, R.O., 1987, Statistical Methods for Environmental Pollution Monitoring. Van Nostrand Reinhold Co., 320 pp. See Chapter 16 & 17 on detecting trends, and Appendix B Testing for Monotonic Trends Using Mann-Kendall, Seasonal Kendall, and Related Nonparametric Techniques.

Harcum, J.B., Loftis, J.C., and Ward, R.C., 1992, Selecting trend tests for water quality series with serial correlation and missing values: Water Resources Bull., V. 28, No.3, p. 469-478.

Heitler, W.J., 2004, Robust Fit, A computer program for robust smoothing of data using locally-weighted scatterplot smoothing (LOWESS), or robust fitting of data to a polynomial. http://www.st-andrews.ac.uk/~wjh/robustfit/

Helsel, D.R. and Cohn, T.A., 1988, Estimation of descriptive statistics for multiply censored water quality data: Water Resources Research, V. 24, No. 12, pp. 1997-2004.

Helsel, D.R. and Hirsch, R.M., 2002, Statistical Methods in Water Resources: U. S. Geological Survey Techniques of Water-Resources Investigations, Book 4, Hydrologic Analysis and Interpretation, Chapter A3, 510 pp.

Hirsch, R.M., Slack, J.R., and Smith, R.A., 1982, Techniques of trend analysis for monthly water quality data: Water Resources Research, V. 18, No. 1, pp. 107-121.

Hirsch, R.M. and Slack, J.R., 1984, A nonparametric trend test for seasonal data with serial dependence: Water Resources Research, V. 20, No. 6, pp. 727-732.

Hirsch, R.M., Alexander, R.B., and Smith, R.A., 1991, Selection of methods for the detection and estimation of trends in water quality: Water Resources Research, V. 27, No. 5, pp. 803-813.

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

65

Hollander, M. and Wolfe, D.A., 1973, Nonparametric Statistical Methods. John Wiley & Sons. See Table A.21 upper tail probabilities for the null distribution of Kendall's K statistic.

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

IDT (Intelligent Decision Technologies, Inc.), 1998, WQStat Plus User's Guide. This software supports Seasonal Kendall, Mann-Kendall, Sen's Slope test, and seasonality testing via Kruskal-Wallis.

Kendall, M.G., 1975, Rank Correlation Methods, 4th ed., Charles Griffin, London.

K-H, 2000, Final Rocky Flats Cleanup Agreement, Attachment 5, action levels and Standards for Surface Water, Ground Water, and Soils, dated March 21, 2000.

K-H, 2002, Annual Report for the RFETS Groundwater Plume Treatment Systems, January through December 2001.

K-H, 2004a, Final 2002 Annual RFCA Groundwater Monitoring Report for RFETS. February 2004.

K-H, 2004b, Evaluation of Natural Attenuation and Biodegradation Potential of Chlorinated Aliphatic Hydrocarbon Compounds in Groundwater at Rocky Flats. Report 04-RF-00358, March 24, 2004.

Mann, H.B., 1945, Non-parametric test against trend: Econometrica, V. 13, p. 245-259.

Martin, L.M., 2000, *De facto* data analysis methods for goal oriented monitoring: what does current practice tell us?: Proceedings National Water Quality Monitoring Council. http://www.nwqmc.org/2000proceeding/papers/pap_martin.pdf

Montgomery, D.C., and Peck, E.A., 1982, Introduction to Linear Regression Analysis. John Wiley, NY.

Neter, J., Wasserman, W., and Whitmore, G.A., 1988, Applied Statistics, 3rd Ed., Allyn and Bacon, Inc. 1006 pp. See Classical Time Series Model pp. 820-845.

Reckhow, K.H., Kepford, K. and Hicks, W.W., 1993, Statistical Methods for the Analysis of Lake Water Quality Trends. EPA841-R-93-003.

Regnier, R., 1998, Trend analysis of annual spring-overturn total phosphorus in 8 small lakes in southern Vancouver Island, British Columbia: BC Ministry of Environment, Lands, and Parks. http://wlapwww.gov.bc.ca/wat/wq/trendstuff/vanisle/springoverturn.html

Richerson, P.M., 2003, Northern Malheur County Groundwater Management Area Trend Analysis Report. Oregon Department of Environmental Quality.

Rong, Y., 2002, Laboratory detection limits: Soil Sediment & Water, AEHS Magazine, April/May 2002. http://www.aehsmag.com/issues/2002/april_may/laboratory_detection.htm

86

Rosen, M.R., 2003, Trends in Nitrate and Dissolved Solids Concentrations in Groundwater, Carson Valley, Douglas County, Nevada 1985-2001, USGS WRI 03-4152.

Rosner, B., 1975, On the detection of many outliers: Technometrics, V. 17, p. 221-227.

Schertz, T.L., Alexander, R.B., and Ohe, D.J., 1991, The Computer Program Estimate Trend (ESTREND), A System for the Detection of Trends in Water Quality Data: Water Resources Investigations Report 91-4040, 63pp.

Sen, P.K., 1968, Estimates of the regression coefficient based on Kendall's tau: J. American Statistical Assoc., V. 63, p. 1379-1389.

Smith, D.G. and McCann, P.B., 2000, Water quality trend detection in the presence of changes in analytical laboratory protocols. In Proceedings of the 2000 National Water Quality Monitoring Council. http://www.nwqmc.org/2000proceeding/papers/pap_smith(b).pdf

USGS, 2004a, Availability of S-ESTREND, an S-PLUS based version of ESTREND. http://water.usgs.gov/admin/memo/QW/qw02.12.html

USGS, 2004b, USGS Library for S-Plus for Windows. Code and datasets downloadable at http://water.usgs.gov/software/library.html

87

RESPONSE TO COMMENTS ON DRAFT STATISTICAL METHODS FOR TRENDING GROUNDWATER QUALITY DATA
Dated July 2004

During August 2004, Rocky Flats issued a draft of the above report for review. CDPHE responded with review comments by Edgar Ethington and Elizabeth Pottorff. EPA did not provide written comments. This letter contains our responses to the CDPHE review comments and describes how this report has been finalized.

Comment 1

I concur with the recommendation of the report to use the Seasonal-Kendall test to estimate trends. It is a robust, non-parametric statistical test with a minimum of pre-test data manipulation.

Reply 1

We agree.

Comment 2

I do not concur that semi-annual sample intervals are always preferred. Sample intervals need to be area specific to dynamic hydraulic systems.

Reply 2

Groundwater velocities at RFETS often vary seasonally and by area and lithology of the porous medium. Groundwater velocities on steep hillsides are generally faster. However, the mean groundwater velocity through Rocky Flats Alluvium is only 105 feet per year, and 138 feet per year in Colluvium (2003 RFCA Annual Groundwater Monitoring Report). Since 1996 the routine groundwater monitoring network at RFETS has successfully used semi-annual groundwater sampling, with the exception of RCRA wells, which are on quarterly monitoring.

Comment 3

I recommend a section be added to briefly describe, in mathematical notation, the use and appreciation of the three trending methods described. §3.7 perhaps. R. D. Gibbons (1994) gives a good example of a brief mathematical description.

Reply 3

We don't believe that adding a new section is necessary. Readers who desire a detailed understanding of these methods are referred to the published scientific literature cited in Section 6, References.

88

Comment 4

The assumption that trends will be monotonic is a best guess. I do not disagree that this assumption is an appropriate first estimate. However, other assumptions will be considered during periodic review of the data.

Reply 4

We agree. The data are likely to display a number of fluctuations, up and down. Whether these represent long-term trends might be arguable, but they will likely result from various causes as noted in the trending discussion.

Comment 5

The trending considered in this document is for long-term multi-year trends.

Reply 5

We concur. A number of years of data are required to identify significant trends.

Comment 6

§2.5 Extra effort needs to be made to minimize lost data points. If samples are lost or rendered unusable, the sample should be retaken as soon a practicable.

Reply 6

Replacement of lost or unusable water samples is a reasonable idea and should be considered in the procedures for post-closure groundwater monitoring.

Comment 7

§2.6 Last paragraph. In a seasonal data set with more than one sampling event, a seasonal average result is preferable to a randomly chosen datum.

Reply 7

The report follows the findings of a paper by statisticians associated with Colorado State University, who indicate that collapsing data by random subsampling is statistically more effective than using the mean or median. See Harcum, Loftis, and Ward, 1992, Selecting trend tests for water quality series with serial correlation and missing values: Water Resources Bulletin, V. 28, No. 3, p. 469-478.

Comment 8

§2.10 The Department concurs that trends in water quality data are often non-linear. Changes in water quality more often exhibit an exponential form.

89

Reply 8

We agree that water quality trends are often non-linear; standard industry practice is to test for monotonic trends but such data should not be assumed linear.

Comment 9

§2.10 Last paragraph. Another objective of post-closure monitoring will be to demonstrate ground water goals are being achieved, or not.

Reply 9

We concur. The outcome of a trending analysis could show conditions other than those anticipated or desired. Contingencies should be planned for such eventualities.

Comment 10

§2.11 Second paragraph. I do not concur with a suggested proposal to include data records for wet and dry years in the initial trend analysis. Site closure represents the start of a new hydraulic equilibrium. Wet and dry year results from a former equilibrium will skew the results.

Reply 10

We concur that closure represents the start of a new hydraulic equilibrium. However, a minimum of four years of water quality data will likely be required before trend testing can usefully be employed. Thus, if stakeholders desire trend testing to start following the first year of post-closure monitoring, it will be necessary to base the trend on at least three years of pre-closure data. If trend testing were to begin with the first CERCLA five-year review, then no pre-closure data need be used.

Comment 11

§3.2 Use of exogenous variables in trend analysis may be useful. It may not.

Reply 11

We concur. The selection of an appropriate exogenous variable, and an accurate and quantifiable understanding of its influence, is critical to its successful application to a trending analysis. Our initial application of an exogenous variable failed to enhance previously identified trends. We do not propose applying trending analysis compensated by such variables.

Comment 12

§3.3 Manipulation of data should be kept to a minimum. Since the Seasonal-Kendall method does not manipulate or smooth data, it is preferred to a method that computes a seasonal mean or grand mean as part of the testing process.

90

Reply 12

We agree; that is part of the basis for our proposal.

Comment 13

§4.4.2 I would not make the initial assumption of normality for water data sets, particularly contaminant data sets. How do the data sets look when non-normality is the null assumption?

Reply 13

The Seasonal-Kendall method has been selected because it does not require any assumption or proof of normality. We agree that many water quality datasets are not normally distributed. The application of parametric methods requires that the data were randomly drawn from a normal distribution, and eliminates such methods from further consideration. Section 4.4.2 simply reports the results of testing the groundwater data for normality using two standard statistical methods, the Shapiro-Wilk and Shapiro-Francia tests. The null hypothesis of each test is that the data are normally distributed. By design, hypothesis tests seek sufficient evidence to reject the null hypothesis at a given level of confidence, otherwise the null hypothesis is accepted. We are unaware of a statistical test for non-normality, i.e. one in which the null hypothesis is that the data are not normally distributed.

Comment 14

§4.4.4 I am not convinced that auto-correlation of quarterly monitoring data is a problem that should preclude its use. It appears that auto-correlation of semi-annual and quarterly data is approximately the same. In other words, auto-correlation is not sufficient reason for the facility to preclude quarterly sampling.

Reply 14

We agree that autocorrelation should not preclude the use of quarterly groundwater monitoring data. RFETS routinely uses quarterly data collected at RCRA wells. Evidence of serial correlation in the semiannual datasets was only slightly less frequent than it was in the quarterly datasets. However, that being said, the historical record at Rocky Flats does not suggest that the well concentrations are changing at a rate that would be perceptible sooner should the data be collected more frequently. Even semi-annual sampling may be more frequent than would be minimally acceptable to reliably detect significant trends.

Comment 15

Figure 4-13 – Well P115689 looks like a new problem, did we keep it in the monitoring network?

Reply 15

VOCs such as PCE, TCE, and VC were first detected in groundwater at Well P115689 during 1993 and the well has undergone many years of water quality monitoring. Well P115689 and nearby Well P115589 both contain VOC-contaminated groundwater of the IA plume near B-551. Well P115589 remains in the monitoring network, but P115689 is no longer included in the

91

network. Because this is the central-IA, numerous downgradient wells monitor the progression of this plume.

Comment 16

P4-44 unfinished sentence in paragraph below bullets.

Reply 16

The sentence has been completed.

92/92

Admin Record Master Entry (FISF_AR_ADMIN_RECORD)

CERCLA Administrative Record Database

Activity	Level	Doc. No	Doc Date	Est Pages	Routine	Status
SW	A	005020	11/10/2004	92	YES, ROUTINE	PRELIM

Print

Title: Statistical Methods for Trending Groundwater Quality Data

Internal Code: 04-RF-01164; (Ref: 04-DOE-00886; 00583-RF-04) Rev No Tag

Document Type: REPORT / STUDY / PLAN

Date Received: 12/02/2004

Login Date: 12/16/2004 Login By: JCR

Data Entry Date: 12/16/2004 Indexer: JCR

Receipt Type Under Review

Prelim History Public History Core A/R Document?

Create Date: 12/16/2004

By User: N902699

Last Modified: 09/27/2005

By User: N711573

Title/Subject Acronym

Statistical Methods for Trending Groundwater Quality Data, November 10, 2004. This evaluation has reviewed much of the published literature regarding trend analysis methods. Emphasis was placed on trend analysis of environmental water quality data.

Comments Acronym

1 CD attached to document.