ATTACHMENT 5
STATISTICAL ANALYSIS OF GROUNDWATER MONITORING DATA

1.0 INTRODUCTION

As part of the post-closure permit for the Hazardous Waste Impoundments (HWI) at the Geneva Steel site in Vineyard, Utah, the Utah Department of Environmental Quality, Division of Solid and Hazardous Waste (UDEQ/DSHW) requires post-closure monitoring of groundwater the uppermost aquifer beneath the HWI. The requirement includes statistical evaluation of groundwater monitoring data collected from wells within and adjacent to the HWI. This document, *Attachment 5, Statistical Analysis of Groundwater Monitoring Data,* describes the statistical methods used to evaluate the groundwater data.

Modules III and IV of the post-closure permit require a monitoring program to detect the release to groundwater of hazardous constituents from the HWI and to ensure compliance with the groundwater protection standards referenced in the permit. The 13 monitoring wells listed in Exhibit 1 and shown graphically in Exhibit 2 constitute the network of monitoring wells specified in Modules III and IV of the permit. Exhibit 1 also indicates whether a well is a background or compliance well, and the depth group to which each monitoring well is assigned; the segregation into groups by depths is necessary for proper application of the statistical analysis methodology described below. Note that some wells belong to more than one group because of the length of the screened interval in the wells. Sampling of the well network is performed in accordance with the schedule specified in Table 1 of Attachment 6 of the permit, and to all other requirements of the permit. Statistical analysis of the groundwater monitoring data is performed after each sampling event.

Exhibit 3 graphically depicts the statistical methodology used to evaluate the data collected from the HWI monitoring network. The methodology is consistent with the requirements outlined in R315-8-6, *Standards for Owners and Operators of Hazardous Waste Facilities, Groundwater Protection* of the Utah Administrative Rules. It includes computation of the descriptive statistics for each measured groundwater constituent for each monitoring well, time series plots for selected constituents, evaluation of whether the data for a constituent can be considered to be a sample drawn from a normal distribution, and hypothesis testing procedures (analysis of variance, the *t*-test, the Kruskal Wallis test and the Wilcoxon Rank-Sum test) used to compare data from compliance wells to data from background wells. The remainder of *Attachment 5* describes the methodology in more detail, beginning with a discussion of how nondetect samples are treated. Brief explanations of hypothesis testing and parametric and nonparametric statistical procedures follow. The decision logic depicted in Exhibit 3 is then thoroughly described, particularly the application of the various hypothesis testing procedures.

In addition to monitoring of groundwater in the uppermost aquifer beneath the HWI, as specified in Modules III and IV of the post-closure permit, Module V of the permit requires perimeter groundwater monitoring for the entire Geneva Steel site. The network of perimeter monitoring wells and piezometers comprises the locations listed in Exhibit 4 and depicted graphically in Exhibit 5. Included in Exhibits 5 and 6 are sentry wells located interior to the site that will be evaluated using by the same statistical methods used to evaluate the perimeter wells. The statistical analysis methodology applied to the groundwater data derived from the perimeter monitoring network and the sentry wells is summarized in Exhibit 6.

Defining background groundwater concentrations for the constituents detected in the perimeter and sentry locations is problematic because many of the wells located on the eastern property boundary, and thus upgradient in terms of the horizontal component of the hydraulic gradient, show elevated levels of some groundwater constituents. For example, dissolved arsenic was measured at a concentration of 310 micrograms per liter (μg/L) in a groundwater sample taken from PZ-01D, near the northeastern corner of the site, during December 2004. This is only slightly below the concentration of 320 μg/L measured in a sample taken from MW-109M, on the western boundary of the site, during the same sample event. The difficulty in defining background is due to two primary factors: (1) naturally-occurring spatial variability in groundwater chemistry, and (2) the presence of industrial facilities to the east, and upgradient, of the Geneva Steel site. Because of the difficulty defining in background conditions, an intrawell approach will be used to evaluate groundwater data derived from the perimeter and sentry locations. In an intrawell approach the only data used in the statistical evaluation of a well are from the well itself. The intrawell analysis includes computation of the descriptive statistics for each measured groundwater constituent for each monitoring well, time series plots for selected constituents, calculation of the lower confidence limit (LCL) for the median based on the seven most recent measurements and comparison of the median with the site-specific screening level (SSSL) for the constituent, time series plots of median derived from the seven most recent samples, and estimates of trend based on the entire historical record for the constituent in the well. Note that the analyte list for the monitoring and sentry locations is considerably longer than for the HWI monitoring wells because it consists of the groundwater monitoring list in 40 CFR 264 Appendix IX.

Compared to the HWI wells, the perimeter and sentry locations have a short historical sampling record, which limits the number of statistical procedures that can be applied to the data. Future modification of the methods applied to data acquired from the perimeter and sentry locations may be warranted as new data become available and understanding of groundwater chemistry improves.

2.0 NONDETECTS

Many of the dissolved groundwater constituents sampled at the HWI often occur at a level below the laboratory method detection limit (MDL) for the constituent. The data for the constituent are therefore censored, because the true concentration of the constituent cannot be estimated due to laboratory limitations. For example, toluene has been measured above the laboratory detection limit in only five of the 25 samples taken from MW-21 between April 1990 and October 2003, and during the same period toluene has never been measured above the detection limit in MW-22, MW-23, and MW-24. A common practice employed in the statistical analyses of censored groundwater data is substitution, i.e., to replace nondetects with an arbitrary numerical value, typically one-half the detection limit reported by the laboratory.

This approach is increasingly recognized as inappropriate, particularly when a large (> 15%) fraction of the measurements are below the detection limit, because it introduces bias in the calculation of sample statistics and the estimation of population parameters, distorts the sample histogram by introducing a spike (or spikes, in the case of multiple detection limits) at one-half the MDL, and negatively affects the results of parametric hypothesis testing (EPA, 1992; Helsel and Hirsch, 2002; Helsel, 2005). Two steps are taken to mitigate the effects of nondetects on the estimation of population parameters and hypothesis testing. First, population parameters such as the mean and standard deviation are estimated using techniques that account for nondetects without resorting to substitution (Helsel, 2005). These techniques include the Kaplan-Meier and maximum likelihood methods (Helsel, 2005). Second, nonparametric methods are preferred when making comparisons between upgradient and downgradient wells. Nonparametric methods do not require an assumption about a true underlying data distribution (which can be difficult to infer with censored data), using instead the relative positions (ranks) of the data rather than the reported numerical values (Conover, 1999).

3.0 HYPOTHESIS TESTING

The comparison between background and compliance wells relies on statistical hypothesis testing, a method of inferring from sampled data whether or not a given statement about one or more populations is true (Conover, 1999). Testing is performed using two hypotheses, the null hypothesis H_0 and the alternative hypothesis H_A. In groundwater monitoring, the two hypotheses take the following general form for a particular dissolved constituent:

H_0: There is no difference in concentration between the background and compliance wells

H_A: The compliance wells have higher concentrations than the background wells

The relative likelihood of the two hypotheses is evaluated using an appropriate test statistic computed from the sampled data. The value of the test statistic is an indicator of whether to accept or reject the null hypothesis (Ostle and Mensing, 1975). Following the example above, one possible test statistic is the difference between the median of the concentration measurements from the compliance well(s) and the median of the measurements from the background well(s). If the difference is small, it would seem reasonable to accept the null hypothesis and conclude that there is no difference between the background and compliance wells. Conversely, if the median computed from the compliance well(s) is much larger than the median for the background well(s), it would be reasonable to reject the null hypothesis in favor of the alternative, and conclude that groundwater samples from the compliance wells might be affected by site contaminants.

Probability provides the means to quantify the concepts of "small," "large," and "reasonable," and also to specify the risk of error. Occurrence probabilities for possible values of the chosen test statistic are determined either by computation or from tables. If the null hypothesis is true, and there is in fact no difference in concentration between the background and compliance wells, then in the example above the probability is small that a large difference in median concentrations will occur. There is always a risk, however, that a large difference in observed medians is a random occurrence, due to the use of sampled data rather than to actual (and unknown) differences in the underlying populations, in which case the null hypothesis is erroneously rejected. This type of error, falsely rejecting the null hypothesis when it is true, is called *Type I* error (also called a *false positive*). By design, the probability of Type I error, denoted by α, is made small, commonly 0.05 or 0.01. Thus, when the computed value of the test statistic has an occurrence probability of less than α, H_0 is rejected. The probability α is also called the significance level of the test, and the quantity $1-\alpha$ is called the confidence level of the test.

A second type of error, failing to reject the null hypothesis when it is false, is called *Type II* error; the probability of committing a Type II error is denoted by β. Unlike α, β is not specified, but depends on the true (and unknown) value of the test statistic as determined from the true (unknown) underlying populations. In the example above, if the true background and compliance medians differ, but by only a small amount, the probability β will be high; if the true difference in medians is large, however, then the probability β will be small.

4.0 PARAMETRIC vs. NONPARAMETRIC METHODS

Parametric statistical methods assume that the sample data come from population with a known distribution (Conover, 1999). The most common assumption is that the sample data are drawn from a normal distribution, or a closely related distribution such as the lognormal. Verification of this assumption requires accepting the null hypothesis of goodness-of-fit tests such as the Shapiro Wilk *W* test or D'Agostino test (Gilbert, 1987). As noted in Section 2.0, however, the presence of nondetects distorts the sample distribution, negatively affecting the reliability of conclusions inferred from these tests (Helsel, 2005).

Nonparametric methods, also known as distribution-free methods, require no assumptions about the population probability distribution (Conover, 1999). These procedures use data ranks or sample quantiles, and therefore take advantage of the information conveyed by nondetects in the form of population proportions. Nonparametric methods are often equally effective to parametric methods, and in some cases are superior, particularly with positively-skewed data and in the presence of data outliers, common situations encountered in environmental investigations (Conover, 1999). Substitution of numeric values for the nondetects is not necessary and there is no need to transform sample data by taking logarithms or powers of the data in order to apply nonparametric methods.

5.0 STATISTICAL ANALYSIS METHODOLOGY, HWI MONITORING WELLS

The analytical results for past data obtained from the HWI wells have been reported using at least two distinct reporting protocols. Data from the earliest sampling events were reported using laboratory reporting limits (RLs), whereas data from subsequent events were reported using MDLs. Unfortunately, some of the RLs in the early data are often an order of magnitude or more greater than either actual sample values or the MDLs. This discrepancy severely limits the utility of the early data because the high RLs provide information of negligible beneficial value. For example, the RL in 1999 for arsenic in MW-10 (a shallow background well) was 100 µg/L, whereas in subsequent events, measured concentrations were less than 11 µg/L. Clearly, little valuable statistical information can be derived from RLs that are greater than the estimated range of concentrations. In these cases, data reported as less than excessively high RLs are excluded from statistical computations.

The majority of the statistical computations will be performed using the commercial NCSS® software package available from Number Crunching Statistical Systems (NCSS [Hintze, 2004]). NCSS® is a comprehensive statistical software that includes nearly all of the techniques described in this document, including the Kaplan-Meier method. MDL (Helsel, 2005) is an additional program that may be used that computes descriptive statistics for a sample in the presence of nondetects.

Statistical methods can only be applied when an adequate number of data are available. This is true both theoretically (larger sample sizes yield more reliable statistics) and practically (numerical algorithms perform poorly when sample sizes are small). Accordingly, the parametric statistical analysis will be limited to those constituents and locations for which (1) at least five samples are available, and (2) with a proportion of nondetects of no more than 50%. Non-parametric analysis will be performed on all data sets.

5.1 DESCRIPTIVE STATISTICS AND TIME SERIES PLOTS

The initial step in the methodology presented in Exhibit 3 is the calculation of the descriptive statistics and the construction of time series plots for selected constituents in each well. The wells and constituents are listed in Table 1 of Attachment 6 of the HWI post-closure permit.

For each constituent in a given well, the descriptive statistics are computed using the entire available historical record of measurements except data omitted based on excessively high RLs or MDLs. The descriptive statistics computed for each constituent fall into two groups. The first group comprises the sample median, quartiles, interquartile range, minimum, maximum, absolute range, and the total number of samples and proportion of nondetects; the second group comprises the sample mean and the standard deviation. The reason for dividing the sample statistics into two groups stems from the impact of nondetects on the computation of the statistics. Nondetects do not affect the first group of sample statistics, provided that the statistics are reported in a way that preserves the information carried by the nondetects. For example, in the simple case where the entire data sample for a constituent in a well consists of nondetects, each of the statistics (except for the number of samples and proportion of nondetects) in the first group would be equal to <MDL, where MDL is the method detection limit for the constituent and, for the sake of illustration, is assumed constant throughout the historical record of measurements. In a somewhat less obvious example suppose the 25^{th} and 75^{th} percentiles of a data set are, respectively, <MDL and $c_{0.75}$; then the interquartile range (*IQR)* would be reported as $<(c_{0.75}$-MDL). The key point is that no substitution is necessary for nondetects when computing the descriptive statistics of the type in the first group. In contrast, calculation of the sample statistics in the second group requires an approach to account for data that occur below the MDL. Either the Kaplan-Meier method (Helsel, 2005) or the maximum likelihood method (Helsel, 2005) will be employed to compute the sample mean and standard deviation in the presence of nondetects.

Time series plots of the historical record for a constituent in a well permit visual evaluation of temporal trends in measured concentrations. Each plot will display the permitted concentration limit (PCL) specified in the post-closure permit, if a PCL exists. Differing symbols will be used to distinguish between detections and nondetects. Nondetects will be assigned a value of zero solely for the purpose of constructing the plots. Time series plots will be generated only for those constituents and locations for which (1) have at least five samples available, and (2) with a proportion of nondetects of no more than 50%.

5.2 PROPORTION OF NONDETECTS

In addition to the descriptive statistics, the proportion of nondetects for each constituent in each compliance well will be computed. Because of practical considerations, a data sample set must have a proportion of nondetects of less than 50% for additional data analysis.

5.3 GOODNESS-OF-FIT TESTING FOR NORMALITY

Each constituent in a well will be tested for normality using the Shapiro-Wilk *W* goodness-of-fit test (Gilbert, 1987; EPA, 1992). The *W* test will be performed to evaluate whether the data set can be considered to be a sample drawn from a normally distributed population, and thus to determine whether parametric or nonparametric methodologies are appropriate in subsequent comparisons between compliance and background wells. In order to perform the test, nondetects will be dropped from the sample set. Unfortunately, this compromise for dealing nondetects is deemed a necessary practical tradeoff in order to accomplish goodness-of-fit testing. If a suitable technique for performing goodness-of-fit testing in the presence of nondetects is identified it will be proposed to UDEQ/DSHW. As noted in Section 2.0, when a large proportion of the data are nondetects, such as often occurs with data from the HWI wells, goodness-of-fit testing becomes problematic when data substitution is used. Thus, experience with data from HWI wells indicates that although normality may be inferred for a constituent in a particular well, it cannot be inferred for the constituent in all wells within a well group. This observation holds whether the *W* test is performed using the raw data or the logarithm of the data. The equivocal results of goodness-of-fit testing, largely a consequence of the typically high proportion of nondetects, reinforce the selection of nonparametric methods for interwell comparisons.

The results of the goodness-of-fit testing for normality will be used to determine whether parametric or non-parametric methods are used to compare concentrations in compliance wells with background.

5.4 COMPARISONS WITH BACKGROUND

The method used to compare the concentrations of a groundwater constituent measured in samples from compliance wells against concentrations measured in samples from a background well (or wells) depends on the number of compliance wells involved in the comparison and whether the data sample is consistent with the characteristics of a normal distribution. When multiple compliance wells are compared to background (either a single well or the pooled result from multiple wells), either parametric analysis of variance (ANOVA) methods or the non-parametric Kruskal-Wallis test is used. When a single compliance well is compared to a single background well, either the parametric *t*-test for means (a special case of the ANOVA method) or the Wilcoxon Rank-Sum test is used. These well-established methods are described in EPA (1992) and Conover (1999). Conover (1999) provides the better and more comprehensive description of the non-parametric methods, including exhaustive examples and a readable discussion of theory.

5.4.1 PARAMETRIC METHODS

Data analyzed using the parametric methods described below must satisfy the requirements that (1) each of the data samples can be considered as a sample drawn from a normal distribution; and (2) the (unknown) variance of each of the underlying normal distributions is equal (Ostle and Mensing, 1975). Thus to evaluate whether a particular constituent in the compliance wells is elevated with respect to background, the sample data from *all* wells must be a sample from some normal distribution, and the underlying normal distributions must *all* have the same variance. Parametric methods will be used if both of these requirements are satisfied.

It is unlikely, however, the *W* tests for normality will indicate that all well-specific samples for a constituent are consistent with a normal population, particularly when the proportion of nondetects in a well is high. It is also unlikely the sample variance computed from the data from each well will indicate that population variances are equal. (The modified Levene test of variance was used to evaluate the hypothesis of equal variance in NCSS [Hintze, 2004].) Accordingly, the comparisons between compliance and background wells will probably be performed using non-parametric methods. Nonetheless, discussion of the parametric methods is included for completeness.

5.4.1.1 ANALYSIS OF VARIANCE (ANOVA)

One-way ANOVA is a parametric method for comparing constituent concentrations in multiple compliance wells to background (either a single well or the pooled result from multiple background wells). The test assumes that (1) all samples are random samples from their respective normal population, (2) the variance of the normal populations are identical, and (3) in addition to independence within each sample, there is mutual independence among the various samples. The test will be performed at the α = 0.01 significance level. The null and alternative hypotheses are:

> H_0: The means of each of the underlying normal populations are identical
>
> H_A: Not all of the means are equal

Thus, rejecting the null hypothesis suggests that at least one well exhibits higher constituent concentrations than in other wells. Failing to reject indicates no difference between compliance wells and background. Note that rejecting H_0 is not evidence of contamination. For example, all concentrations in a well may be below the PCL, but at a level that is slightly elevated in comparison to other wells. Moreover, the test makes no distinction between background and compliance wells; they are compared as a group, so elevated concentrations in the background well may lead to rejecting H_0. Finally, the result applies to the group of wells tested, not to an individual well. In the event of rejection, evaluation of individual wells is necessary, including comparison to the constituent PCL, time series plots, prediction intervals, and evaluation of the hydrogeological setting of the well. Clear explanations of the computational details for the one-way ANOVA method test are presented in EPA (1992) and Ostle and Mensing (1975).

5.4.1.2 *t*-TEST

The two-sample *t*-test is a parametric test for comparing the means of two samples (Ostle and Mensing, 1975). The test is used to compare a single well against background. The test assumes that (1) both samples are random samples from normal populations, (2) the variances of the underlying populations are identical and (3) in addition to independence within each sample, there is mutual independence among the various samples. The test will be performed at the α = 0.01 significance level. The null and alternative hypotheses are:

> H_0: Both underlying populations have the same mean
>
> H_A: The mean of the compliance well population is higher than that of the background population

The alternative hypothesis above is for an upper-tailed test, and is one of three statements of H_A that are possible with the test (see Ostle and Mensing, 1975). Rejecting the null hypothesis suggests that the downgradient well exhibits higher constituent concentrations than the background well. Failing to reject indicates no difference between the compliance well and background. Rejecting H_0 is not evidence of contamination, since all concentrations in a well may be below the PCL, but at a level that is slightly elevated in comparison to the background well. In the event of rejection, evaluation of the compliance well is necessary, including comparison to the constituent PCL, time series plots, prediction intervals, and evaluation of the hydrogeological setting of the well. Clear explanations of the computational details for the *t*-test are presented in EPA (1992) and Ostle and Mensing (1975).

5.4.2 NON-PARAMETRIC METHODS

Non-parametric methods are more applicable to the HWI because of the very small likelihood that the data sets used in the evaluations will all meet the requirements for parametric testing discussed above.

5.4.2.1 KRUSKAL-WALLIS TEST

The Kruskal-Wallis test is a nonparametric counterpart to the parametric ANOVA method, and is used to compare constituent concentrations in multiple compliance wells to background (either a single well or the pooled result from multiple background wells). The test assumes that (1) all samples are random samples from their respective populations, and (2) in addition to independence within each sample, there is mutual independence among the various samples. The test will be performed at the $\alpha = 0.01$ significance level. The null and alternative hypotheses are:

> H_0: The population distribution functions for each sample are identical
>
> H_A: At least one of the populations tends to yield larger observations than at least one of the other populations

H_A is sometimes recast to state that the populations do not have equal means. Thus, rejecting the null hypothesis suggests that at least one well exhibits higher constituent concentrations than in other wells. Failing to reject indicates no difference between compliance wells and background. Note that rejecting H_0 is not evidence of contamination. For example, all concentrations in a well may be below the PCL, but at a level that is slightly elevated in comparison to other wells. Moreover, the test makes no distinction between background and compliance wells; they are compared as a group, so elevated concentrations in the background well may lead to rejecting H_0. Finally, the result applies to the group of wells tested, not to an individual well. In the event of rejection, evaluation of individual wells is necessary, including comparison to the constituent PCL, time series plots, prediction intervals, and evaluation of the hydrogeological setting of the well. Clear explanations of the computational details for the Kruskal-Wallis test are presented in EPA (1992) and Conover (1999).

5.4.2.2 WILCOXON RANK-SUM TEST

Also known as the Mann-Whitney test, the Wilcoxon Rank-Sum test is a nonparametric counterpart to the parametric two-sample *t*-test for comparison of means (Conover, 1999). The test is used to compare a single well against background. The test assumes that (1) both samples are random samples from their respective populations, and (2) in addition to independence within each sample, there is mutual independence among the various samples. The test will be performed at the $\alpha = 0.01$ significance level. The null and alternative hypotheses are:

H_0: The population distribution functions for each sample are identical

H_A: The expected value from one population is greater than the expected from the other population

The alternative hypothesis above is for an upper-tailed test, and is one of three statements of H_A that are possible with the test (see Conover, 1999). Rejecting the null hypothesis suggests that the downgradient well exhibits higher constituent concentrations than the background well. Failing to reject indicates no difference between the compliance well and background. Rejecting H_0 is not evidence of contamination, since all concentrations in a well may be below the PCL, but at a level that is slightly elevated in comparison to the background well. In the event of rejection, evaluation of the compliance well is necessary, including comparison to the constituent PCL, time series plots, prediction intervals, and evaluation of the hydrogeological setting of the well. Clear explanations of the computational details for the Wilcoxon Rank-Sum test are presented in EPA (1992) and Conover (1999). (Note: Conover (1999) denotes the test as the Mann-Whitney test.)

5.5 PREDICTION INTERVALS (INTRAWELL HISTORY COMPARISON)

Additional evaluation is warranted when either of the Kruskal-Wallis or Wilcoxon Rank-Sum tests suggests that a constituent concentration may be elevated in a compliance well. The first step in the evaluation is to compare recently measured concentrations against the historical record of concentrations in the well. The historical record is divided into a background period, consisting of a specified number of past measurements, and the prediction period, consisting of the measurements more recent than the period. In this sense, the prediction limit is not a bound on actual future concentrations, but a hypothetical bound on concentrations measured after some fixed time in the past. As few as one sample (the most recently measured concentration) can constitute the prediction period. This intrawell comparison is accomplished by estimating the nonparametric upper prediction limit (EPA, 1992), which is simply the maximum value of the measurements in the background period.

The prediction limit has confidence probability 1-α that the next future sample or samples will be below the prediction limit. The confidence 1-α depends on the number of samples for the constituent used for the prediction period, denoted by n, and the number of future samples k: $(1-\alpha) = n/(n+k)$, or $\alpha = 1 - n/(n+k)$. Thus, the significance level α is not specified, but is computed. Given a fixed number of available samples $n+k$, the only way to decrease α is to decrease the number of predicted samples and increase the number of background samples. For example, if $n+k = 24$ and $k = 8$, then $\alpha = 0.33$, an excessively high false positive rate. Decreasing k to one yields $\alpha = 0.04$. When intrawell prediction limit calculations are warranted, k will be set equal to one. Note that in the future the significance level α will continually decrease as the number of sampling events increase.

Exceedance of the intrawell prediction limit suggests the potential of groundwater contamination due to past activities at the site. As discussed in the next section, however, elevated constituent concentrations in a well may be due to hydrogeological conditions, which must be considered before concluding that groundwater contamination has occurred.

5.6 HYDROGEOLOGICAL SETTING

The Geneva Steel site is located on the eastern shoreline of Utah Lake, which exerts major hydrogeological influence on subsurface flow at the site because it is a regional groundwater sink. Groundwater mainly derived from precipitation in the Wasatch Mountains, located approximately 8 miles east of the Geneva Steel site, ultimately discharges to Utah Lake. As a result, the local hydraulic gradient at the site yields groundwater flow that is generally directed upward and to the west. In addition, groundwater originating in the mountains carries with it dissolved constituents due to mineralization in the mountains, modified by the subsurface geochemical conditions encountered as the groundwater moves from the mountains to the lake.

In this setting, the deep HWI monitoring wells can be considered to provide background information for some of the groundwater constituents measured in samples from the shallower wells. Arsenic, for example, is ubiquitous in the deep monitoring wells as a group, generally occurring at concentrations four to ten times the PCL and above the concentrations measured in shallow wells. Based on review of time series plots, this pattern appears to be temporally persistent. It is unlikely that the elevated levels of arsenic in the deep monitoring wells are due to site impacts because the upward component of flow would tend to prevent dissolved contaminants from migrating downward. Thus, any evaluation of arsenic in groundwater at the site should consider the spatial distribution of arsenic in the subsurface before concluding that contamination has occurred in deep compliance wells.

The example of arsenic described above is only one example of how natural conditions can affect the distribution of dissolved constituents in groundwater. Other natural hydrogeological circumstances, unanticipated and not explored here, may also affect groundwater chemistry at the HWI, and should be evaluated before concluding that groundwater contamination has occurred due to the HWI.

6.0 ANALYSIS METHODOLOGY, PERIMETER AND SENTRY MONITORING LOCATIONS

In contrast to the HWI monitoring wells, the perimeter monitoring locations are intended to monitor groundwater quality at the facility boundary. The sentry wells are intended to delineate impacts to groundwater quality due to a specific solid waste management unit (SWMU) or group of SWMUs (SWMUG). The risk-based SSSLs will be used with both the perimeter and sentry wells as a benchmark to identify potentially compromised groundwater quality.

Fewer sampling events are available for the perimeter and sentry locations compared with the HWI monitoring network. This is the main reason for using different statistical procedures to evaluate data from the perimeter and sentry locations. When sufficient data become available in the future (approximately 15 total sampling events), review of the statistical methods is warranted. At that time, the same general procedures used to evaluate data from the HWI network may be applicable to data from the perimeter and sentry locations.

Similar to the HWI wells, the analytical results for past data obtained from the perimeter and sentry wells have been reported using at least two distinct reporting protocols for the perimeter monitoring locations. Data from the earliest sampling events were reported using laboratory RLs, whereas data from subsequent events were reported using MDLs. Unfortunately, the RLs of the early data are often several orders of magnitude greater than either actual sample values or the MDLs. This discrepancy severely limits the utility of the early data because the high RLs provide information of negligible beneficial value. For example, the RL in 1997 for 1,2-dichloroethane in MW-100S (a perimeter well) was 5,000 μg/L, whereas in subsequent events, measured concentrations were on the order of 1 μg/L or less or below a MDL of 0.26 μg/L. Clearly, little valuable statistical information can be derived from RLs that are greater than the estimated range of concentrations. A similar situation occurs when the MDL applied to early sampling events are greater than reported values from subsequent events. In these cases, data reported as less than excessively high RLs or MDLs are excluded from statistical computations.

The majority of the statistical computations will be performed using the commercial NCSS® software package available from NCSS (Hintze, 2004). NCSS® is comprehensive statistical software that includes nearly all of the techniques described in this document, including the Kaplan-Meier method. Two additional programs may be used: Trend Version 2.01, documented in Gilbert (1987) performs the Mann-Kendall test for trend and Sen's estimate of slope; and MDL (Helsel, 2005) computes descriptive statistics for a sample in the presence of nondetects.

Statistical methods can only be applied when an adequate number of data are available. This is true both theoretically (larger sample sizes yield more reliable statistics) and practically (numerical algorithms perform poorly when sample sizes are small). Accordingly, the parametric statistical analysis will be limited to those constituents and locations for which (1) at least five samples are available, and (2) with a proportion of nondetects of no more than 50%. Non-parametric analysis will be performed on all data sets.

6.1 DESCRIPTIVE STATISTICS AND TIME SERIES PLOTS

Descriptive statistics and time series plots will be generated as described in section 5.1. Exceedance of the respective SSSL by any constituent will be noted.

6.2 PROPORTION OF NON-DETECTS

In addition to the descriptive statistics, the proportion of nondetects for each constituent in each compliance well will be computed. Because of practical considerations, a data sample set must have a proportion of nondetects of less than 50% for additional data analysis.

6.3 EVALUATE DATA TRENDS (MANN-KENDALL METHOD)

The Mann-Kendall test is a non-parametric method that can be used for evaluating whether the data exhibit a temporal trend. It is particularly appropriate for use in the PGMP because irregularly spaced data and nondetects can be accommodated (Gilbert, 1987). The test assumes sample independence. The test will be performed at the $\alpha = 0.05$ significance level. The null and alternative hypotheses are:

H_0: The data do not exhibit a trend

H_A: The data exhibit an upward trend

The alternative hypothesis above is for an upper-tailed test, and is one of three statements of H_A that are possible with the test (see Gilbert, 1987). Rejecting the null hypothesis suggests a trend of increasing concentration for the constituent in the well. Gilbert (1987) and Gibbons (1994) provide discussion and examples of the Mann-Kendall test, including the necessary probability tables for the test statistic. As much of the historical record of data as possible will be used to evaluate the presence of an upward trend (see discussions regarding RLs in Section 6.0).

6.4 LOWER CONFIDENCE LIMIT FOR THE MEDIAN

This approach is analogous to using the four most recent samples to compute the lower confidence limit for the mean and using the computed lower limit to infer exceedance of some stipulated groundwater standard. The lower confidence limit for the median is used in lieu of the more common lower confidence limit for the mean because of the typically high proportion of nondetects that occur in groundwater samples from the Geneva Steel site and because a reasonably low probability α of Type I error can be achieved with seven samples. In addition, the median is a better measure of central tendency than the mean for positively-skewed data (Ostle and Mensing, 1975), such as often occurs with groundwater monitoring data.

The lower one-sided confidence limit for the median will be used to determine exceedance of the SSSL for each constituent in each well and will be computed using the seven most recent samples. When seven samples are not available, the maximum possible odd number of samples will be used; for example, if six data are available, the five most recent data will be used. When seven data are available, the lower limit for the 94% confidence interval ($\alpha = 0.06$) for the median is given by the second of the ordered data, i.e., the next to the smallest (minimum) of the data (Conover, 1999). Note that the size of the confidence interval is fixed for a given number of data, and decreases when fewer samples are used in the calculation. When fewer than seven data are available, Conover (1999) provides a table (Table A.3 – Binomial Distribution) for evaluating the size confidence interval (in each case, the second of the ordered data will be used as the LCL). Exceedance of the SSSL by the lower confidence limit for the median is statistical evidence that the constituent is above its SSSL in groundwater, indicating degradation in groundwater quality with respect to the constituent in question.

6.5 TIME SERIES PLOTS OF THE MEDIAN

As a complement to time series plots of the actual data, a time series plot will be generated of the median computed using the seven most recent data. The plot will also display the LCL for the median, computed as described in Section 6.4, and the SSSL for the constituent. The purpose of the plot is to provide the means to visually evaluate the trend; a persistent increase in the value of the computed median suggests that increasing concentrations of the constituent from upgradient sources are reaching the well. These plots will not be generated until the third round of applying the data analysis described above because the plots will not begin to have any value until at least three data are posted to the plot.

7.0 REFERENCES

Conover, W.J., 1999. *Practical Nonparametric Statistics.* John Wiley and Sons, New York.

Gibbons, R.D., 1994. *Statistical Methods for Groundwater Monitoring.* John Wiley and Sons, New York.

Gilbert, R.O., 1987. *Statistical Methods for Environmental Pollution Monitoring.* John Wiley and Sons, New York.

Helsel, D.R., 2005. *Nondetects and Data Analysis: Statistics for Censored Environmental Data.* John Wiley and Sons, New York.

Helsel, D.R. and R.M. Hirsch, 2002. *Statistical Methods in Water Resources.* U.S. Geological Survey Techniques of Water Resources Investigations, Book 4, Chapter A3. U.S.

Hintze, J., 2004. Number Crunching Statistical Systems (NCSS). Kaysville, Utah. www.NCSS.com.

Ostle, B. and R.W. Mensing, 1975. *Statistics in Research.* The Iowa State University Press, Ames, IA.

U.S. Environmental Protection Agency (EPA), 1992. *Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities, Addendum to Interim Final Guidance.* U.S. EPA, Washington, D.C.

Exhibit 1. Hazardous Waste Impoundment Groundwater Monitoring Wells

Shallow Wells		Intermediate Wells		Deep Wells	
Well ID	**Type**	**Well ID**	**Type**	**Well ID**	**Type**
MW-10	Background	MW-1	Background	MW-19D	Background
MW-12	Compliance	MW-5	Compliance	MW-14D	Compliance
MW-21	Compliance	MW-8	Compliance	MW-17D	Compliance
MW-22	Compliance	MW-9	Compliance		
MW-23	Compliance	MW-13	Compliance		
MW-24	Compliance	MW-21	Compliance		
		MW-22	Compliance		
		MW-23	Compliance		
		MW-24	Compliance		

Exhibit 2. Hazardous Waste Impoundment Groundwater Monitoring Wells Site Map

Exhibit 3 Statistical Evaluation Flowchart for the Hazardous Waste Impoundment Wells

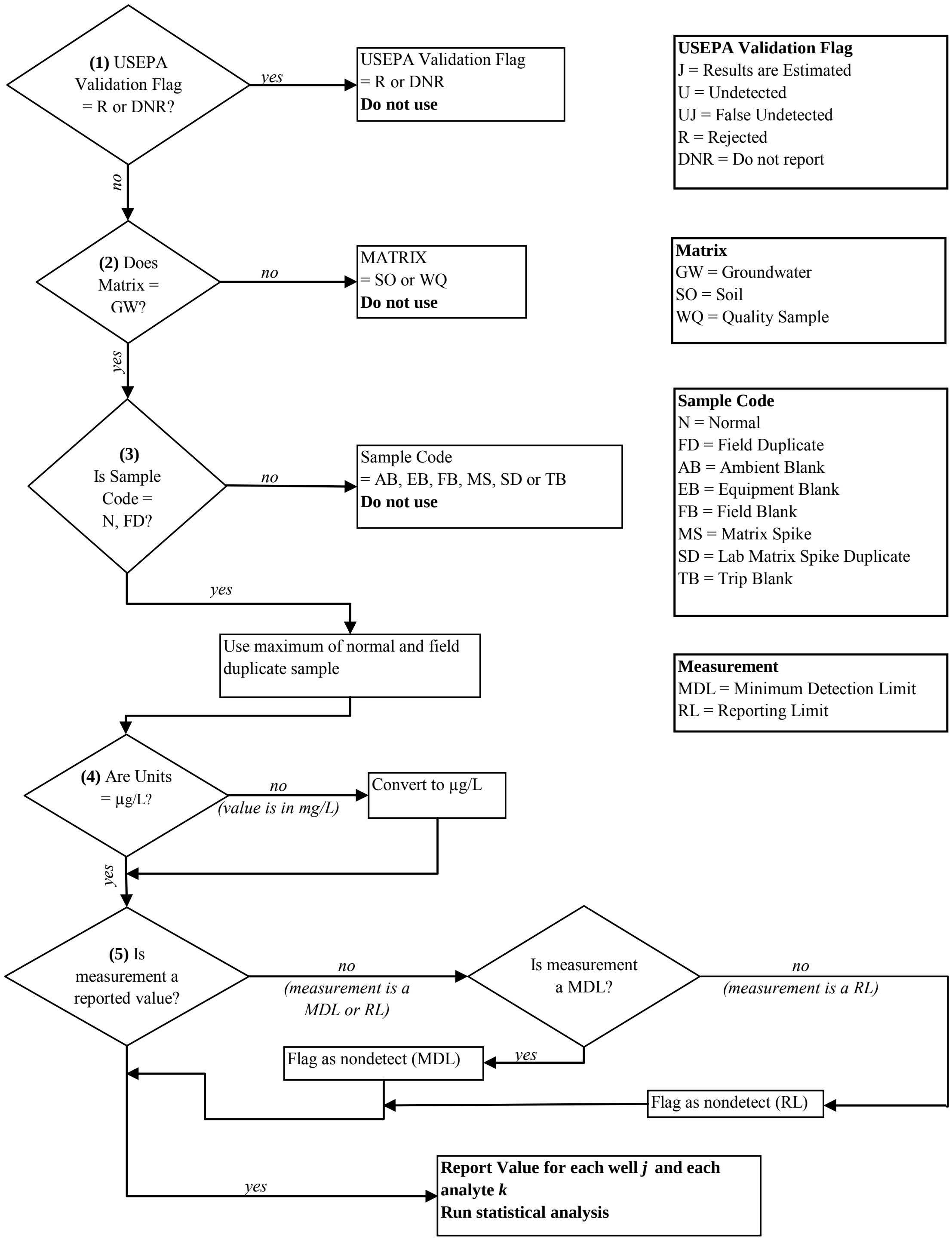

Exhibit 3 (continued)
Statistical Evaluation Flowchart for the Hazardous Waste Impoundment Wells

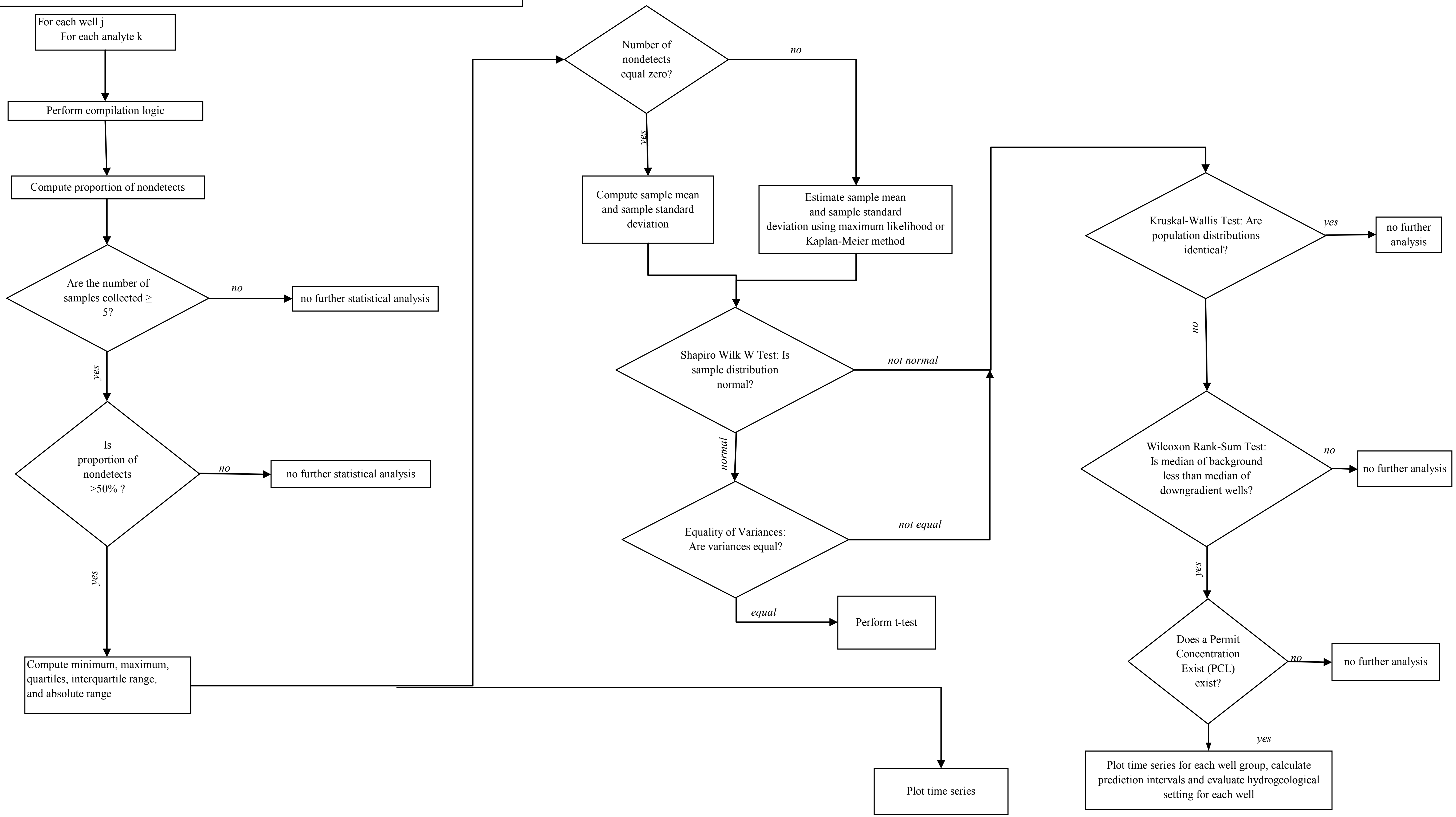

Exhibit 4. Perimeter and Sentry Groundwater Monitoring Locations

Perimeter-In Wells Well ID		
MW-100S	MW-116S	PZ-03M
MW-101S	MW-117S	PZ-03SR
MW-102S	MW-118S	PZ-04D
MW-103S	MW-119S	PZ-04M
MW-104M	MW-120S	PZ-04S
MW-104S	MW-121S	PZ-05S
MW-105S	MW-122S	PZ-06S
MW-106M	MW-123S	PZ-07D
MW-106S	MW-124S	PZ-07S
MW-107S	MW-125S	PZ-08D
MW-108S	MW-126S	PZ-08M
MW-109M	MW-127S	PZ-08S
MW-109S	MW-128S	PZ-09D
MW-110S	MW-129S	PZ-09M
MW-111S	MW-130S	PZ-09S
MW-112M	PZ-01D	PZ-11S
MW-112S	PZ-01M	PZ-12S
MW-113S	PZ-01S	PZ-13D
MW-113M	PZ-02D	PZ-13M
MW-114M	PZ-02M	PZ-13S
MW-114S	PZ-02S	PZL-16
MW-115S	PZ-03D	PZL-16

PZ-10S, on previous lists, has been destroyed
PZL-18, on previous lists, cannot be located
MW-112M and MW-112S were abandoned after the 2006 sampling event
MW-113M was added beginning with the 2006 sampling event

Exhibit 5. Perimeter and Sentry Groundwater Monitoring Locations

Exhibit 6 Statistical Evaluation Flowchart for the Perimeter and Sentry Wells

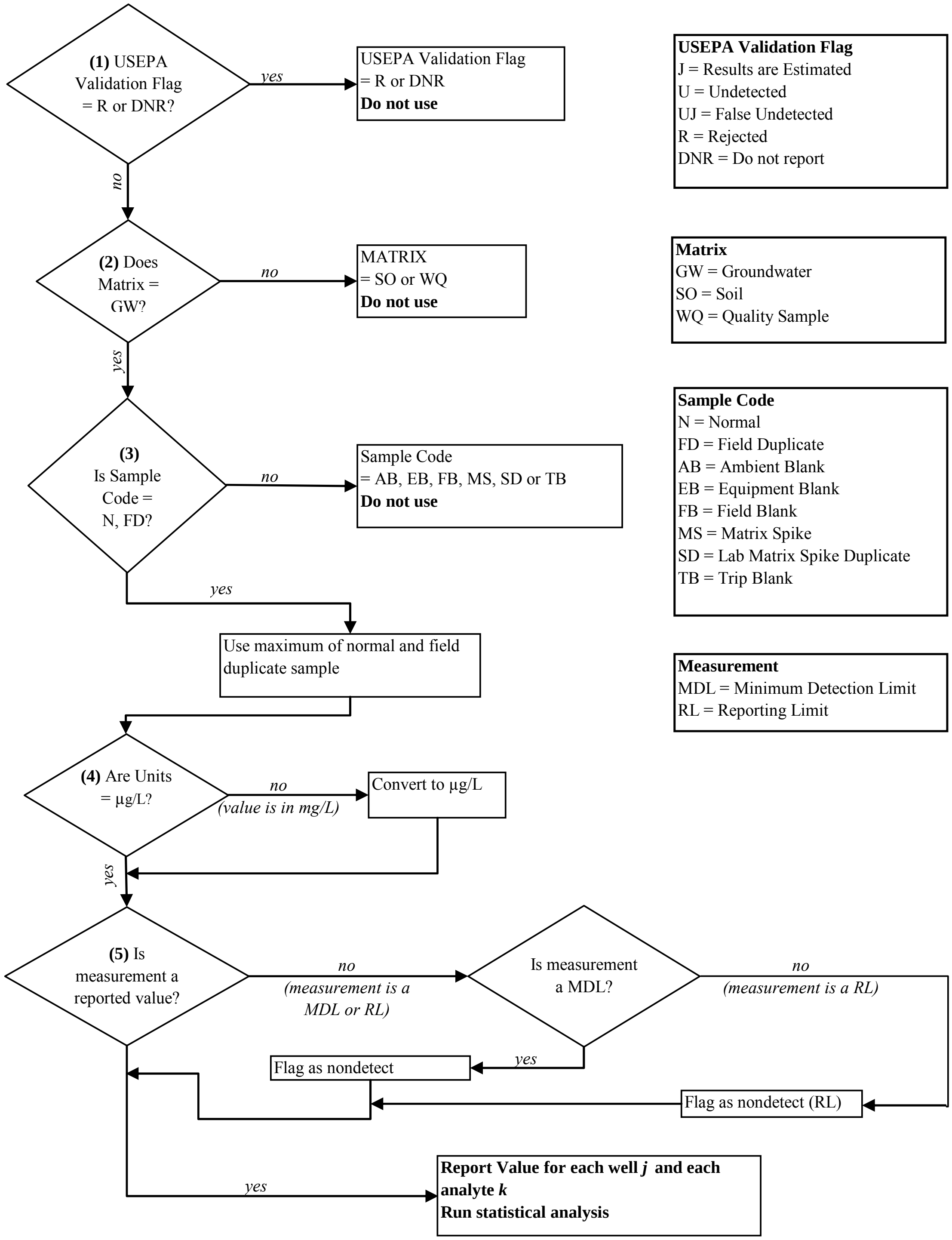

Exhibit 6 (continued) Statistical Evaluation Flowchart for the Perimeter and Sentry Wells