COAL COMBUSTION RESIDUAL RULE
STATISTICAL ANALYSIS PLAN

REVISION NO. 1

MARTIN LAKE STEAM ELECTRIC STATION
PERMANENT DISPOSAL POND 5
RUSK COUNTY, TEXAS

NOVEMBER 16, 2022

Prepared For:

Luminant Generation Company LLC

Prepared By:

Golder Associates USA , Inc.
1601 S. Mopac Expy, Suite 325D
Austin, Texas 78746
Texas Engineering Firm Registration No. 22771

November 16, 2022

PROFESSIONAL CERTIFICATION

This document and all attachments were prepared by Golder Associates USA, Inc. under my direction or supervision. I hereby certify that the proposed statistical method is appropriate for evaluating groundwater data in accordance with the requirements of Sections 257.93 through 257.95 of the CCR Rule.

Patrick J. Behling, P.E.
Principal Engineer
GOLDER ASSOCIATES USA, INC.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ACRONYMS AND ABBREVIATIONS

ANOVA	analysis of variance
CCR	coal combustion residuals
EPA	United States Environmental Protection Agency
GWPS	groundwater protection standard
LCL	lower confidence limit of the mean
MCL	maximum contaminant level
PPCC	Filliben's probability plot correlation coefficient test
RROS	robust regression order statistics
SAP	statistical analysis plan
SWFPR	site-wide false positive rate
UPL	upper prediction limit

1.0 INTRODUCTION

The United States Environmental Protection Agency (EPA) issued regulations regarding the disposal of coal combustion residuals (CCR) in certain landfills and impoundments in April 2015. These regulations, found under 40 CFR 257, Subpart D and referred to as the "CCR Rule" require facilities to design a groundwater monitoring program to monitor if landfills or impoundments with CCR materials, called CCR units, are impacting downgradient groundwater quality.

Section 257.90 of the CCR Rule requires that all existing CCR landfills and surface impoundments comply with the following groundwater monitoring requirements no later than October 17, 2017:

- Install a groundwater monitoring system as required under Section 257.91;
- Develop a groundwater sampling and analysis program to include selection of the statistical procedures to be used for evaluating groundwater monitoring data as required under Section 257.93;
- Initiate a detection monitoring program to include obtaining a minimum of eight independent samples for each background upgradient and downgradient monitoring well as required under Section 257.94; and
- Begin evaluating the groundwater monitoring data for statistically significant increases over background levels for the constituents listed in Appendix III of this part as required under Section 257.94.

Statistical analysis of groundwater monitoring data is required as part of detection monitoring and assessment monitoring under Section 257.93 of the CCR Rule. Section 257.93 of the CCR Rule provides several options for statistically evaluating groundwater data. The owner or operator of the CCR unit must select one of the statistical methods specified in paragraphs (f)(1) through (5) of Section 257.93 when evaluating constituent concentrations from the groundwater monitoring. EPA's *Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities, Unified Guidance* (EPA, 2009), also called the "Unified Guidance", presents acceptable statistical approaches for such evaluations and analyses. However, neither the CCR Rule nor the Unified Guidance outlines a step-by-step process to consistently evaluate groundwater monitoring data in order to satisfy the CCR Rule.

The purpose of this statistical analysis plan (SAP) is to develop a standard set of statistical approaches to follow when demonstrating groundwater compliance for each CCR unit in accordance with the CCR Rule and the Unified Guidance. Depending on the CCR unit and the evaluation of groundwater data for the CCR

unit, CCR groundwater compliance may be evaluated using either an interwell or an intrawell approach—the interwell approach being a comparison of water quality data upgradient of the CCR unit to water quality data downgradient of the CCR unit, and the intrawell approach being a comparison of water quality data of a well against background values established from that well's own historical water quality data.

This SAP describes and summarizes the statistical approach for establishing and evaluating baseline conditions to use for detection monitoring and assessment monitoring. The plan is designed to detect a release from a CCR facility. The plan conforms with EPA "Unified Guidance Document: Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities," March 2009, and the American Society for Testing and Materials (ASTM) Standard D6312-17, Developing Appropriate Statistical Approaches for Groundwater Detection Monitoring Programs at Waste Disposal Facilities.

2.0 DATA PREPARATION

Analytical data from wells in the groundwater monitoring network at a CCR unit during each sampling event are first reviewed for usability after final data packages are received from the laboratory. The analytical data are then prepared for statistical analysis. Methods for handling duplicate and non-detect data are implemented during this data preparation phase in order to comply with the performance standards outlined in 40 CFR 257.93. During the data preparation, anomalously low or high constituent concentrations are also considered for usability. The following subsections provide further details.

2.1 Handling Duplicate Data

Field duplicates and data rejected after data validation are removed from the data set. Only the primary samples are retained for the statistical evaluation.

2.2 Handling Non-Detect Data

A non-detected constituent concentration is defined as any analytical result that either has an instrument response but is below a sample detection limit or that has no instrument response. A non-detected concentration is handled by using one of two approaches, depending on the percentage of detections in the data set:

- If a data set has at least 85% of samples detected, half of the sample detection limit is substituted as a proxy concentration. In these cases, substituting a proxy concentration will not alter the results of statistical tests or summary statistics (EPA, 2009; EPA, 2000).

- If a data set has at least 50% but no more than 85% of the samples detected, the robust regression order statistics (RROS) method is used to estimate summary statistics such as the mean and standard deviation (EPA, 2009).

- If a data set has fewer than 50% of the samples detected, then nonparametric statistical approaches are used to evaluate the data and to prepare summary statistics (EPA, 2009; EPA, 2000).

It should be noted that J-flagged data (estimated concentrations between the sample detection limit and the reporting limit) are defined as detected concentrations.

2.3 Handling Anomalous Detections

There may be infrequent cases when an anomalously high or low detection cannot be confirmed after resampling a well. In such cases, the anomalous detection should be considered for removal from the data

set and should be replaced by the resampled concentration so that current conditions are not over- or under-estimated. This is particularly important when estimating a baseline or background value to use to compare to future constituent concentrations from the network of groundwater monitoring wells. An anomalous detection may be identified at any point after analytical laboratory results are available, based on professional judgment or based on the outlier evaluation (see Section 3.4 for more details about testing for outliers). If an analytical result is removed, documentation should be provided in the annual report stating which analytical result was removed and justifying its removal.

3.0 STATISTICAL ASSUMPTIONS

Before baseline or background values can be established, a number of statistical assumptions are evaluated to determine if concentrations are independent and identically distributed. A sample's constituent concentration is independent when no other sample concentrations influence its measurement, regardless of when or where the sample was collected. Statistical independence is indicated by a set of random data. But randomness is only demonstrated by the presence of mean and variance stationarity and the lack of evidence for effects such as spatial and temporal variation, autocorrelation, and trends (EPA, 2009).

The validity of statistical independence is checked by testing for:

- Spatial stationarity,
- Temporal stationarity,
- Lack of autocorrelation, and
- Lack of statistical data outliers.

For the purpose of this SAP, the statistical software R (The R Foundation, 2017) is assumed to be used to perform the statistical tests used for checking the validity of independent samples. Other applicable programs may be used as necessary.

3.1 Spatial Stationarity

Spatial stationarity is defined as the lack of variability across well locations. Spatial variation may be naturally occurring and unaffected by human activity, or may be caused by human activity. The presence of spatial variability does not necessarily mean that contamination is present. If spatial variability is present, regardless whether it's naturally-occurring or not, it may hinder attempts to identify the cause of a statistically significant increase in constituent concentrations between current and baseline or background conditions (EPA, 2009). In some cases, spatial variability may make upgradient-to-downgradient comparisons (also called interwell comparisons) difficult (EPA, 2009).

One way to identify spatial stationarity is to observe whether spatial variability does or does not exist across multiple wells. This is particularly true when a CCR unit has more than one upgradient well and when interwell comparisons are used for detection or assessment monitoring. Constituent concentrations from each upgradient well are taken as a single data set and then upgradient well data sets are compared. Before establishing baseline or background values for the detection monitoring or assessment monitoring

programs, two steps are taken to check for spatial stationarity for each constituent and groundwater monitoring well (recommended by the Unified Guidance):

1. Side-by-side box plots are created, and
2. The one-way analysis of variance (ANOVA) or Kruskal-Wallis test is used.

Box plots provide a quick screen for possible spatial variation. The ANOVA and Kruskal-Wallis test are more formal tests for identifying spatial variability. All of the statistical tests are performed and the box plots are generated using the statistical software R (The R Foundation, 2017) or similar software.

In some cases, spatial variability, where substantial differences in average constituent concentrations are present among upgradient wells, can make interwell comparisons difficult (EPA, 2009). Professional judgment should be used to determine whether the set of constituent concentrations from all upgradient wells appropriately represent baseline or background conditions and whether the spatial variability will prevent the detection or assessment monitoring from identifying a potential release at a CCR unit. If the spatial variability were to indicate that analytical data from a set of upgradient wells do not appropriately represent background conditions or if the spatial variability were to hinder the detection or assessment monitoring, then the data set should be adjusted accordingly.

3.1.1 Box Plots

A box plot is a graphical representation of the pattern and distribution of concentrations for a single constituent data set. Visually comparing box plots for upgradient well's constituent concentrations, side-by-side, is one way to identify similarities or differences across upgradient well concentrations. If box plots contain similar range of concentrations, then the concentrations for the upgradient wells are similar (spatial stationarity). Likewise, if box plots do not contain similar range of concentrations, then the concentrations for the upgradient wells are different: spatial variability. Section 3.4.1 provides more details about how to create box plots.

3.1.2 ANOVA and Kruskal-Wallis Tests

The ANOVA and Kruskal-Wallis tests are similar statistical tests; both tests indicate significant spatial variability by indicating whether a statistically significant difference exists among average, upgradient well concentrations. The ANOVA is a parametric approach for comparing average concentrations across two

or more wells. The Kruskal-Wallis test is a non-parametric approach to the ANOVA using the ranks of concentrations, rather than using the actual concentration measurements. Neither test can be performed if the variances across upgradient wells are unequal. A Type I error rate (α), or level of significance, is set to α=0.05 for identifying a statistical significant different among well averages.

Determining which test to perform, either the ANOVA or Kruskal-Wallis tests, depends upon the frequency of detected results, the validity of assuming normality or lognormality for residuals, and the validity of assuming upgradient wells have equal variances. More details about these dependencies are provided in the subsections below (Sections 3.1.2.1-3.1.2.3). Figure 1 outlines the steps taken to define which statistical test (ANOVA or Kruskal-Wallis) should be used. The method used to determine the appropriate statistical test is based on the Unified Guidance recommendations. Tests of normality and equal variances use a 0.01 level of significance, rather than a 0.05 level of significance, because the ANOVA is reasonably robust to small departures of normality and equal variances (EPA, 2009).

No statistical test is performed when there are no detected concentration measurements in any of the upgradient wells.

If there are at least 85% detected concentrations in every upgradient well, then the ANOVA may be considered. For any non-detected concentration, half of the sample detection limit is used as a proxy concentration (see Section 2.2 for more details). The assumptions of normality and equal variances are checked. To test the normality assumption, residuals are tested using two distributional tests, the Shapiro-Wilk test and Filliben's probability plot correlation coefficient (PPCC) test. The Levene's test is used to check for equal variances. Only when evidence exists that both assumptions are valid is the ANOVA using the raw concentration measurements used. If either assumption is not met, then the assumptions of normality and equal variances are checked using the log-transformed data. Only when evidence exists that both assumptions are valid is the ANOVA using the log-transformed concentration measurements used. If either assumption is not met, then an ANOVA cannot be considered.

If there are fewer than 85% detected concentrations or if the ANOVA cannot be considered, then the Kruskal-Wallis may be considered. Non-detected data are treated differently for the Kruskal-Wallis test since the ranks of the data are used rather than the concentration measurements: all data below the maximum sample detection limit are set to the same value, lower than the maximum sample detection limit (Helsel, 2012). Since the Kruskal-Wallis tests uses ranks of the data, the actual value used for data below the maximum sample detection limit is not relevant. The assumption of equal variances is checked using

the Fligner's test. If the Fligner's test indicates that the assumption of equal variances is valid, then the Kruskal-Wallis test is used. Otherwise, no test can be performed because variances are heterogeneous among upgradient well concentration measurements.

3.2 Temporal Stationarity

Temporal stationarity is the lack of temporal variability. Temporal variability refers to the concept that concentration measurements vary over time. Temporal variability may be present across a group of wells and/or constituents. Temporal variability can also be present at an individual well or for a single constituent. By definition, temporal variability also includes autocorrelation, which is discussed separately in Section 3.3.

Any temporal pattern can invalidate or weaken the results of statistical testing (EPA, 2009). Plotting concentrations over time for a given constituent and for a given well is one way to identify possible trends. The Mann-Kendall trend test is another way to identify possible temporal variation for a given constituent and well. The Mann-Kendall is a nonparametric method to test for an increasing or decreasing linear trend over time. The Mann-Kendall doesn't require any special treatment for non-detects, other than all non-detects should be set to a common value lower than any of the detected concentrations (EPA, 2009 p.8-32). The Mann-Kendall is performed for any set of data with at least one detected concentration.

Before establishing baseline or background values for the detection monitoring or assessment monitoring programs, two steps are taken to check for temporal stationarity for each constituent and groundwater monitoring well:

1. A time plot is created, and
2. The Mann-Kendall trend test is used.

The time plots are generated and the Mann-Kendall trend test is performed using the statistical software R (The R Foundation, 2017) and the EnvStats package (Package 'EnvStats', 2017) or similar software.

Statistically significant increasing or decreasing temporal trends are not expected for any upgradient well since, by definition, an upgradient well should not be impacted by a release at the CCR unit. If, however, there is evidence of a temporal trend, then professional judgment should be used to determine whether constituent concentrations from that upgradient well appropriately represent baseline or background conditions and whether the trend will prevent the detection or assessment monitoring from identifying a

potential release at a CCR unit. If the trend were to indicate that an upgradient well does not appropriately represent baseline or background conditions or if the trend were to hinder the detection or assessment monitoring, then the data set should be adjusted accordingly.

To identify a statistically significant temporal trend, a Type I experiment wise error rate (α) is set to $\alpha = 0.05$. That means, a single test error rate is defined for each well across the detected Appendix III or Appendix IV constituents. Each well's single test error rate is based on the number of detected constituents, d, for a given constituent list. For example, a well with five detected Appendix IV constituents ($d = 5$) has a single test error rate equal to $1 - (1 - \alpha)^{1/d^*} = 1 - (1 - 0.05)^{1/5} = 0.0102$. A statistically significant linear trend is identified when the p-value for the Mann-Kendall test is less than the single test error rate.

3.3 Lack of Autocorrelation

Autocorrelation is the statistical dependence between pairs of constituent concentrations across a sequence of time. That is, pairs of consecutive concentrations will exhibit stronger similarity in concentration measurements than expected from pairs collected at random times (p.6-25, EPA, 2009). To identify autocorrelation, the Unified Guidance recommends using the rank von Neumann ratio test for its ease of use and robustness when applied to either normal or non-normal distributions (p.14-17 EPA, 2009). Since this test has not been designed to handle tied values such as non-detect concentrations, this test is only performed for those wells and constituents with at least 50% detected concentrations.

The rank von Neumann ratio test statistic and associated p-value are computed using the statistical software R (The R Foundation, 2017) and the EnvStats package (Package 'EnvStats', 2017) or similar software.

Before baseline or background values are established for the detection monitoring or assessment monitoring programs, the rank von Neumann ratio test is used. Statistically significant autocorrelation is not expected for any well since, by definition, constituent concentration measurements from a well should be collected with far enough time between sampling events that a more recent sample does not include the same volume of groundwater as any previous sample. If, however, there is evidence of autocorrelation, then professional judgment should be used to determine whether constituent concentrations from a well appropriately represent baseline or background conditions and whether the trend will prevent the detection or assessment monitoring from identifying a potential release at a CCR unit. If the trend were to indicate that a well does not appropriately represent baseline or background conditions or if the trend were to hinder the detection

or assessment monitoring, then the data set should be adjusted accordingly.

To identify a statistically significant autocorrelation, a Type I experiment wise error rate, α, of 0.05 is used for each well across the detected Appendix III or Appendix IV constituents. Each well's single test error rate is based on the number of constituents detected at least 50% of the time, d^*, for a given constituent list. For example, a well with five detected Appendix IV constituents ($d^* = 5$), has a single test error rate equal to $1 - (1 - \alpha)^{1/d^*} = 1 - (1 - 0.05)^{1/5} = 0.0102$. A statistically significant autocorrelation is identified when the p-value for the rank von Neumann test is less than the single test error rate.

3.4 Lack of Statistical Outliers

Based on the Unified Guidance, outliers are "extreme, unusual-looking measurements". An outlier may be an invalid concentration measurement due to a typographical error, an equipment error, a sampling error, etc. Or an outlier may be a valid concentration measurement that reflects a "...temporary, local 'hot spot' of higher concentration" (EPA, 2009). Furthermore, outliers are "measurements (larger or smaller than other data values) that are not representative of the sample population from which they were drawn" (EPA, 2002).

The Unified Guidance recommends testing for outliers to attempt to determine whether a suspect outlier may have been drawn from the same sample population as the rest of the data. "The basic problem with including statistical outliers in analyzing groundwater data is that they do not come from the same distribution as the other measurements in the sample and so fail the identically distributed presumption of most tests" (EPA, 2009).

The consequences of keeping statistical outliers when developing a baseline or background value may lead to an unreasonably high value that will be unable to identify potential releases at a CCR unit. Professional judgment should be used to determine whether to retain or remove any outlier. The Unified Guidance states that outliers generally should not be removed unless some basis for a likely error or discrepancy can be identified. Possible errors or discrepancies include "...values significantly outside the historical ranges of background data" (EPA, 2009). "The decision to discard an outlier should be based on some scientific or quality assurance basis" (EPA, 2000). "A data point should not be eliminated from the background data set simply because it is the highest value that was observed" (EPA, 2002). EPA recommends "…that all data not known to be in error should be considered valid" (EPA, 1989). Furthermore, "[t]he general rule is that a measurement should never be deleted from a data set solely on the basis of an outlier test" (SWDIV,

1999).

Before baseline or background values are established for the detection monitoring or assessment monitoring programs, two steps are taken to check for suspect outliers for each constituent with at least 50% detected concentrations and at each well or set of upgradient wells:

1. A box plot is created to identify suspect outliers, and
2. The Dixon's test or Rosner's test is used.

Possible, or suspect, outliers are identified using a box plot. The statistical outlier tests, the Dixon's test and Rosner's test, are tests to check whether any suspect outlier is a statistical outlier. The box plots are generated and the Dixon's or Rosner's test is performed using the statistical software R (The R Foundation, 2017) or similar software.

3.4.1 Box Plots

Creating a box plot is a visual technique used to identify suspect outliers. Box plots can also demonstrate the pattern and distribution of constituent concentrations for a data set. The size of the vertical box in a box plot indicates where the middle half of the data fall (i.e., the interquartile range, IQR). Concentration measurements that plot further away from the others indicate suspect outliers; for a box plot, these measurements are called mild or extreme outliers (EPA, 2009).

Box plots are constructed to identify two types of suspect outliers: mild and extreme outliers. Suspect outliers are defined in terms of the IQR, represented by the range of the middle half of the data and indicated by the vertical 'box' in a box plot. The IQR is the difference between the upper quartile and the lower quartile of the data. Mild and extreme outliers are identified for small or large sample detected concentration measurements. A high, mild outlier is any detected concentration that exceeds 1.5 times the IQR, but no more than 3 times the IQR, from the upper quartile. A small, mild outlier is any detected concentration that is below 1.5 times the IQR, but no less than 3 times the IQR, from the lower quartile. A high, extreme outlier is any detected concentration greater than 3 times the IQR from the upper quartile. A low, extreme outlier is any detected concentration less than 3 times the IQR from the lower quartile. EPA, 2009 and EPA, 2017 state that mild and extreme outliers should be considered suspect outliers. Computational details for box plots are found in EPA guidance documents (EPA, 2000; EPA, 2009).

3.4.2 Statistical Outlier Tests

A statistical outlier test, either the Dixon's test or Rosner's test, is performed for each data set having at least one suspect outlier in order to determine if the suspect outlier is also a statistical outlier. For a data set with no more than 25 samples, the Dixon's test is used. For a data set with at least 20 samples, the Rosner's test is used. Dixon's test can only test if one detected concentration (i.e., the minimum or the maximum) is a statistical outlier. The Rosner's test can test if one or more detected concentrations are statistical outliers (EPA, 2000; EPA, 2002; EPA, 2009). Computational details for these outlier tests are outlined in EPA documents (EPA, 2000; EPA, 2009). Based on results from the statistical outlier tests, mild and extreme outliers are classified as statistical outliers.

Both statistical outlier tests assume that the data set with the suspect outlier(s) removed is normally distributed (or lognormally distributed if the data are transformed to the natural-log scale). Section 4.1.2 below discusses how to test distributional assumptions of normality or lognormality.

Any extreme, suspect outlier that is also identified as a statistical outlier is evaluated for possible errors or data discrepancies before a baseline or background value is established. Suspect outliers, including those also classified as statistical outliers, should be reviewed for having possible analytical or other quality errors. Professional judgment should be used to determine whether constituent concentrations defined as suspect or statistical outliers should be removed so that baseline or background conditions are properly represented so that detection or assessment monitoring can identify a potential release at a CCR unit. If an outlier does not represent baseline or background conditions or if the outlier hinders the detection or assessment monitoring, then the data set should be adjusted accordingly.

4.0 STATISTICAL APPROACH FOR DETECTION AND ASSESSMENT MONITORING

Section 257.93 of the CCR rule provides several options for statistically evaluating the groundwater data and the performance standards to follow at CCR facilities. At each CCR unit, upper prediction limits (UPLs) are calculated for each detected constituent to establish baseline or background values. To achieve UPLs with sufficient statistical power, the UPLs are designed to include retesting procedures based on the 1-of-2 approach (one assigned sample and one resample—see Section 4.1.3). Using UPLs is one of the preferred methods for comparing groundwater based on the Unified Guidance (EPA, 2009).

UPLs are computed using baseline or background data. The source of the baseline or background data may differ, depending whether interwell or intrawell comparisons are appropriate. "With interwell tests, background is derived from distinct, initially upgradient background wells" (EPA, 2009). "Future data from each of these compliance wells are then tested against this common background. On the other hand, intrawell background [also called baseline] is derived from and represents historical groundwater conditions in each individual compliance well." (EPA, 2009)

There are several considerations to make when determining whether interwell or intrawell comparisons should be performed. To consider interwell comparisons for a CCR unit, the groundwater monitoring data should meet the statistical assumptions of spatial stationarity, temporal stationarity, lack of autocorrelation, and lack of statistical outliers (see Section 3). Furthermore, the CCR unit should

- have at least one upgradient well,
- have a clearly defined groundwater flow direction without any radial flow, and
- not contain highly variable mine spoil.

If any of these conditions cannot be met or if the statistical assumptions cannot be met, then intrawell comparisons should be considered for a CCR unit. Both Gibbons and EPA's Unified guidance recommend using intrawell analyses when spatial variability exists. Both Gibbons and the Unified Guidance caution that intrawell analyses are appropriate in the absence of contamination. Since a CCR unit may be an existing landfill or impoundment that is now under the CCR rule, there is a possibility that contamination may be present. Professional judgment should be used for such CCR units to determine if contamination is likely present, and to determine which type of comparison is more appropriate.

4.1 Calculating UPLs

UPLs are estimated with constituent concentrations that are independent and identically distributed, as described in Section 3. The set of data used to calculate UPLs are based on constituent concentrations from the eight background sampling events and from either:

- upgradient wells for the CCR unit (for interwell comparisons), or
- individual compliance well (for intrawell comparisons).

UPLs must be calculated using a single-test error rate that accounts for the site-wide false positive rate (SWFPR) associated with all of the detection or assessment monitoring comparisons. The SWFPR is set based on the Unified Guidance recommendations and is discussed in more detail in Section 4.1.1.

After assumptions have been checked and outliers have been identified for the appropriate set of data, the data distribution is defined in accordance with EPA guidance (EPA, 2000; EPA, 2002; EPA, 2009; EPA, 2017; SWDIV, 1998). UPLs are then calculated based on the defined data distribution. Distributions are defined using the methodology outlined in Section 4.1.2, and the UPLs are calculated using the methodology described in Section 4.1.3.

The statistical software R (The R Foundation, 2017) or similar software is used to perform all statistical distribution tests and to calculate UPLs.

4.1.1 Defining Single-test error rate

Based on 40 CFR 257.93 (g)(2) and the Unified Guidance, the cumulative SWFPR or Type I experiment wise error rate for yearly monitoring shall be no more than 0.10. That means, a single test error rate must be considerably lower than 0.10. The single test error rate depends on the number of detected constituents and number of compliance wells evaluated in a CCR unit's monitoring program, defined as:
$1 - (1 - \alpha)^{1/cw}$, where:

- α=0.10, the SWFPR;
- c=the number detected constituents for the monitoring program (the Appendix III constituents for detection monitoring or Appendix IV constituents for assessment monitoring); and
- w=the number of compliance wells at the CCR unit.

Sampling frequency is not included in this single-test error rate because UPL calculations are designed to account for the number of sampling events per year.

4.1.2 Defining a Distribution for Background

The type of UPL calculated is based on a data set's defined distribution. Figure 2 outlines the steps to take to define whether a data set follows a normal, gamma, lognormal, or nonparametric distribution. If there are no detections for a data set, no distribution is defined. For a constituent with fewer than 50% detected concentrations, the distribution is defined as nonparametric (EPA, 2000; EPA, 2009).

For each data set with at least 50% detected concentrations and at least 4 samples, the data's distribution is tested using up to three distributional tests, which include the Shapiro-Wilk test, Kolmogorov-Smirnov test, and PPCC test. A test for the gamma distribution is included because EPA, 2017 generally recommends using summary statistics from a gamma distribution before using statistics from a lognormal distribution when both the gamma and lognormal distributional assumptions are valid. All of these distributional tests are recommended by EPA (EPA, 2000; EPA, 2002; EPA, 2009; EPA, 2017). Each distributional test is performed with only the detected data, which reflects how ProUCL performs distributional tests (EPA, 2017).

The method used to define a distribution, using the largest p-value from all of the appropriate tests and comparing it to a 0.05 level of significance, is designed to follow ProUCL's distributional recommendations. It should be noted that for a data set with fewer than five detected samples, the Kolmogorov-Smirnov test and the PPCC test cannot be performed. And, the Kolmogorov-Smirnov test is not used to test for gamma distributions.

If results from any of these three tests indicate the data are normally distributed (when the largest p-value is greater than 0.05), the distribution is defined as normal. If none of the test results indicate normality, the detected data set is tested for the gamma distribution by running the Shapiro-Wilk and PPCC tests. If either test indicates the data set follow a gamma distribution (when the larger p-value is greater than 0.05), the

distribution is defined as a gamma distribution. If none of the test results indicate a gamma distribution, the data set is tested for lognormality by running the Shapiro-Wilk, Kolmogorov-Smirnov, and PPCC tests with the log-transformed detected data. If results from any of these tests indicate the data set is lognormally distributed (when the largest p-value is greater than 0.05), the distribution is defined as lognormal. If none of the distributional test results indicate normality, a gamma distribution, or lognormality, the data's distribution is defined as nonparametric.

4.1.3 Calculating UPLs

UPLs are calculated using a 1-of-2 retesting strategy to ensure comparisons are statistically powerful and to minimize the SWFPR. A 1-of-2 retesting strategy means that if one or more constituent concentrations in a compliance well are above their respective background concentration, a resample is collected to validate or invalidate the background concentration exceedance. According to the Unified Guidance, "A *1-of-m* retesting plan implies that up to *m* groundwater measurements may have to be collected at each compliance well, including the initial observation and *(m-1)* possible resamples. For the test to be valid, all of these sample measurements need to be statistically independent" (EPA, 2009). An independent resample may be collected between sampling events if necessary.

The Unified Guidance defines when a well is in-compliance and out-of-compliance: "If the initial groundwater observation is in-bounds [in compliance with the designed standard], the test is complete and no resamples need to be collected. Only when the first concentration exceeds the UPL, does additional sampling come into play" (EPA, 2009). If all *m* samples (the initial sample plus *m-1* resamples) exceed, then the well is considered out-of-compliance. If none of the *(m-1)* resamples exceed after the initial sample exceeded, then the well can still be considered to be in-compliance (EPA, 2009).

The type of UPL computed (e.g., parametric or nonparametric) is based on the detection frequency and the defined data distribution for each data set, as described in Section 4.1.2. For a constituent with no detected concentration measurements in the baseline or background data, the UPL is set to the reporting limit (EPA, 2009). For a constituent with at least 50% detections, the UPL calculation adjusts for non-detected concentration(s) as described in Section 2.2, and the appropriate UPL calculation is used based on results from the distributional tests. If no parametric distribution (normal, lognormal, or gamma) can be defined for a data set, then a nonparametric UPL is estimated. Since J-flagged data are defined as detected, a calculated UPL may be less than the reporting limit; in such cases, the UPL is set to the reporting limit.

4.2 Establishing Background Values

Background values used for detection monitoring or assessment monitoring are based on UPLs. For detection monitoring (Appendix III constituents), background values are defined as the higher of the UPL and reporting limit. For assessment monitoring (Appendix IV constituents) background values are defined as the highest of the maximum concentration level (MCL), UPL, reporting limit, or other accepted screening level for constituents without MCLs. The reporting limit is included so that a constituent having an UPL below the reporting limit does not have an unfair limitation because most or all of the baseline or background constituent concentrations are below the reporting limit. For each CCR unit, tables of statistically-derived background values will be prepared for each Appendix III and Appendix IV constituent. For interwell comparisons, background values will be developed using upgradient well data. For intrawell comparisons, background values will be developed for each monitoring well using historical data from the well.

4.3 Updating Background Values

As detection or assessment monitoring continues, it is recommended to update baseline or background data sets periodically with valid monitoring concentrations that are representative of groundwater unimpacted by leakage from the CCR unit. The Unified Guidance recommends reviewing and possibly updating background values when enough new concentrations have been collected to perform statistical comparisons. That means, background values should be reviewed about every two or three years during. Failure to update background will exclude factors such as natural temporal variation, changes in field or laboratory methodologies, and changes in the water table due to meteorological conditions or other influences.

5.0 DETECTION MONITORING DATA EVALUATION

Detection monitoring will be performed at each CCR unit's groundwater monitoring system on a semi-annual basis during the active life of the CCR unit and during the post-closure period. Each CCR monitoring well will be sampled for the following Appendix III constituents as part of the detection monitoring program:

- Boron
- Calcium
- Chloride
- Fluoride
- field-measured pH
- Sulfate
- Total Dissolved Solids (TDS)

After every detection monitoring event, the constituent concentrations from each well will be compared to the background values, as described in Section 3 of this plan, to ascertain if a statistically significant increase above background exists. Possible outcomes from comparing the detection monitoring constituent concentrations in each well to their respective background values are as follows:

- All detection monitoring constituent concentrations in a compliance well are less than or equal to their respective background values; or

- One or more detection monitoring constituent concentrations in a compliance well are above their respective background values.

5.1 No Statistically Significant Increase over Background Values

Baseline and background UPLs are based on a 1-of-2 resampling approach, meaning that if zero or one concentration measurements from a series of two independent samples collected from a well do not exceed the appropriate UPL, then a statistically significant increase over baseline or background has not occurred at a CCR unit. This conclusion will be reached if the data indicate either of the following:

- All detection monitoring constituent concentrations in a compliance well are less than or equal to their respective background values; or

- At least one detection monitoring constituent concentration in a well is above the respective background value. If this occurs, the well or wells with constituent concentration(s) above the background value(s) will be resampled and analyzed for the detection monitoring constituent(s) with exceedances. If the resample indicates that the target detection monitoring constituent concentration(s) in the well or wells is less than or equal to their respective background value(s),

then it can be concluded that a statistically significant increase over background for all detection monitoring constituents has not occurred, since concentrations in one sample of the two independent samples do not exceed the appropriate baseline or background value(s).

If the groundwater monitoring data indicates that a statistically significant increase over background has not occurred at the CCR wells, then detection monitoring at all CCR wells will continue on a semi-annual basis.

5.2 Statistically Significant Increase over Background Values

If one or more detection monitoring constituent concentrations in any well is above the respective background value in both the original detection monitoring sample and the resample, then a statistically significant increase over background for the target detection monitoring constituents can be concluded. If a statistically significant increase is indicated, within 90 days Luminant will:

- Establish an assessment monitoring program as described in this plan, or
- Demonstrate that a source other than the CCR unit caused the statistically significant increase over the baseline or background value for a constituent, or that the statistically significant increase resulted from error in sampling, analysis, statistical evaluation, or natural variation in groundwater quality. If a successful demonstration is completed within the 90-day period, the owner or operator of the CCR unit may continue with the detection monitoring program.

6.0 ASSESSMENT MONITORING DATA EVALUATION

Assessment monitoring will be performed at a CCR unit's groundwater monitoring system after a statistically significant increase over background values has been confirmed in that well for one or more of the detection monitoring constituents. Within 90 days of triggering the assessment monitoring program, and annually thereafter, each CCR monitoring well requiring assessment monitoring will be sampled for the following Appendix IV parameters as part of the assessment monitoring program:

- Antimony
- Arsenic
- Barium
- Beryllium
- Cadmium
- Chromium
- Cobalt
- Fluoride
- Lead
- Lithium
- Mercury
- Molybdenum
- Selenium
- Thallium
- Radium 226 and 228 combined

Within 90 days of obtaining the results from the initial assessment monitoring sampling event, all wells in a CCR unit's groundwater monitoring system will be resampled and analyzed for:

- All Appendix III detection monitoring parameters; and
- The Appendix IV assessment monitoring parameters that were detected as part of the assessment monitoring event.

This monitoring will be performed on at least an annual basis thereafter, unless Luminant can demonstrate the need for an alternative monitoring frequency for repeated sampling and analysis for these constituents during the active life and the post-closure care period based on the availability of groundwater.

Within 90 days of obtaining the results from the initial assessment monitoring sampling event, a GWPS will be established for each of the Appendix IV assessment monitoring constituents that were detected in the groundwater monitoring system wells as follows:

- For constituents for which an MCL has been established, the highest of the MCL, UPL, and reporting limit for that constituent; or

- For constituents for which an MCL has not been established, the higher of the UPL, reporting limit, or levels that are equivalent to specified regional screening level (RSL) for that constituent (note: future revisions to the Rule may allow additional flexibility in establishing GWPS for states with EPA-approved CCR permit programs for Appendix IV constituents that do not have a MCL).

Each assessment monitoring constituent will be evaluated to ascertain if a statistically significant increase above the GWPS exists. Possible outcomes are as follows:

- All averages from assessment monitoring constituent concentrations at a well are not statistically greater than to their respective GWPS; or
- One or more averages from assessment monitoring constituent concentrations at a well are statistically greater than their respective GWPS.

6.1 Calculating LCLs

For each assessment monitoring constituent, the 95% lower confidence limit of the mean (LCL) is estimated. The set of data used to calculate LCLs are based on the constituent concentrations from the current year's sampling events and enough previous sampling events to reasonably estimate each LCL (the goal should be to have around eight to ten samples).

LCLs are calculated based on the defined data distribution. The data distribution is defined in accordance with EPA guidance (EPA, 2000; EPA, 2002; EPA, 2009; EPA, 2017; SWDIV, 1998). Distributions are defined using the methodology outlined in Section 6.1.1. The LCLs are calculated using the methodology described in Section 6.1.2.

The statistical software R (The R Foundation, 2017) or similar software is used to perform all statistical distribution tests and to calculate LCLs.

6.1.1 Defining a Distribution for LCLs

The type of LCL calculated is based on a data set's defined distribution. The same methodology for defining a distribution for background, described in Section 4.1.2 and outlined in Figure 2, is used to define the distribution for each assessment monitoring constituent data set as normal, gamma, lognormal, or nonparametric.

6.1.2 Calculating LCLs

The type of LCL computed (e.g., parametric or nonparametric) is based on the detection frequency and the defined data distribution for each data set, as described in Section 6.1.1. For a constituent with no detected concentration measurements, the LCL is set to the reporting limit (EPA, 2009). For a constituent with at least 50% detections, the LCL calculation adjusts for non-detected concentration(s) as described in Section 2.2, and the appropriate LCL calculation is used based on results from the distributional tests. If no parametric distribution (normal, lognormal, or gamma) can be defined for a data set or there are fewer than 50% detections, then a nonparametric, approximate 95% lower confidence limit of the median is estimated.

6.2 No Statistically Significant Increase Over GWPS

A statistically significant increase over the groundwater protection standard has not occurred at a CCR unit when the LCL for every assessment monitoring constituent at a well is less than or equal to the appropriate GWPS.

Assessment monitoring will continue on an annual basis. If for two consecutive assessment monitoring sampling events, the constituent concentrations for all Appendix III constituents are at or below background values and all Appendix IV constituents are shown to be statistically at or below their appropriate GWPS, then assessment monitoring will be terminated and detection monitoring as described in this plan will resume. If the constituent concentrations of any Appendix III constituents are shown to be statistically above background values, but all Appendix IV constituents have no statistically significant increase over their respective GWPS, then assessment monitoring will continue.

6.3 Statistically Significant Increase Over GWPS

A statistically significant increase over the groundwater protection standard has occurred at a CCR unit when the LCL for at least one assessment monitoring constituent at a well is greater than the appropriate GWPS. If a statistically significant increase over groundwater protection standards for any Appendix IV assessment monitoring constituent is confirmed, within 90 days of the initial assessment monitoring event, Luminant will either:

- Initiate an assessment of corrective measures for the CCR unit in accordance with CCR Rule Section 257.96; or
- Demonstrate that a source other than the CCR unit caused the contamination, or that the statistically

> significant increase resulted from error in sampling, analysis, statistical evaluation, or natural variation in groundwater quality. If a successful demonstration is made, the owner or operator must continue assessment monitoring. If a successful demonstration has not been made at the end of the 90 day period, the owner or operator of the CCR unit must initiate an assessment of corrective measures for the CCR unit.

If one or more Appendix IV assessment monitoring constituent concentrations are statistically above the respective groundwater protection standards, and if a source other than the CCR unit cannot be demonstrated to have caused the contamination, a release from the CCR unit is likely and the nature and extent of the release will be further characterized as follows:

- Install additional monitoring wells necessary to define the contaminant plume(s);
- Collect data on the nature and estimated quantity of material released including specific information on the Appendix IV assessment monitoring constituents and the levels at which they are present in the material released;
- Install at least one additional monitoring well at the facility boundary in the direction of contaminant migration and sample this well for all Appendix III detection monitoring parameters and for those Appendix IV assessment monitoring constituents that have been detected as part of assessment monitoring. This monitoring must be performed on at least an annual basis thereafter.
- Sample all CCR unit wells for all Appendix III detection monitoring parameters and for those Appendix IV assessment monitoring constituents that have been detected as part of assessment monitoring. This monitoring must be performed on at least an annual basis thereafter.

7.0 REPORTING REQUIREMENTS

The results of the CCR groundwater monitoring program performed at each CCR unit will be reported yearly in an Annual Groundwater Monitoring and Corrective Action Report. A separate annual report for each CCR unit will document the status of the groundwater monitoring and corrective action program, summarize key actions completed, describe any problems encountered, discuss actions to resolve the problems, and project key activities for the upcoming year. At a minimum, the Annual Groundwater Monitoring and Corrective Action Report will contain the following information:

- A map, aerial image, or diagram showing the CCR unit and all background (or upgradient) and downgradient monitoring wells, to include the well identification numbers, that are part of the groundwater monitoring program for the CCR unit;
- Identification of any monitoring wells that were installed or decommissioned during the preceding year, along with a narrative description of why those actions were taken;
- In addition to all the monitoring data obtained under CCR Rule Sections 257.90 through 257.98, a summary including the number of groundwater samples that were collected for analysis for each background and downgradient well, the dates the samples were collected, and whether the sample was required by the detection monitoring or assessment monitoring programs, as well as the basis for the background values and the statistical methods employed to establish the background values;
- A narrative discussion of any transition between monitoring programs (e.g., the date and circumstances for transitioning from detection monitoring to assessment monitoring in addition to identifying the constituent(s) detected at a statistically significant increase over background levels); and
- Other information required to be included in the annual report as specified in CCR Rule Sections 257.90 through 257.98.

The Groundwater Monitoring and Corrective Action Report for the 2017 monitoring program must be placed in each facility operating record no later than January 31, 2018. Subsequent reports must be placed in the facility operating records no later than January 31 of the year following completion of the groundwater monitoring program from the preceding calendar year. The reports must also be posted to the owner or operator's CCR Rule Compliance Data and Information internet site within 30 days of placing the reports in the operating record.

8.0 REFERENCES

ASTM, 2017. Standard Guide for Developing Appropriate Statistical Approaches for Groundwater Detection Monitoring Programs at Waste Disposal Facilities - D6312-17.

EnvStats, 2016. https://cran.r-project.org/web/packages/EnvStats/EnvStats.pdf.

EPA, 2000. Guidance for Data Quality Assessment Practical Methods for Data Analysis, EPA QA/F-9 QA00 UPDATE. . EPA/600/R-96/084. July.

EPA, 2002. Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA Sites. EPA 540-R-01-003. September.

EPA, 2009. Unified Guidance Document: Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities, EPA 530-R-09-007, March.

EPA, 2017. ProUCL Version 5.1 User Guide, https://www.epa.gov/sites/production/files/2016-05/documents/proucl_5.1_user-guide.pdf . February 1.

Helsel, Dennis R. Statistics for Censored Environmental Data Using Minitab and R, second edition. John Wiley & Sons, Inc., 2012.

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/ , January 26, 2017.

Package 'EnvStats', 2017. https://cran.r-project.org/web/packages/EnvStats/EnvStats.pdf.

SWDIV and EFA West of Naval Facilities Engineering Command, 1998. Procedural Guidance for Statistically Analyzing Environmental Background Data. September.

SWDIV and EFA West of Naval Facilities Engineering Command, 1999. Handbook for Statistical Analysis of Environmental Background Data. July.

The R Foundation, 2017. "R" Software Suite. https://www.r-project.org. February.

FIGURES

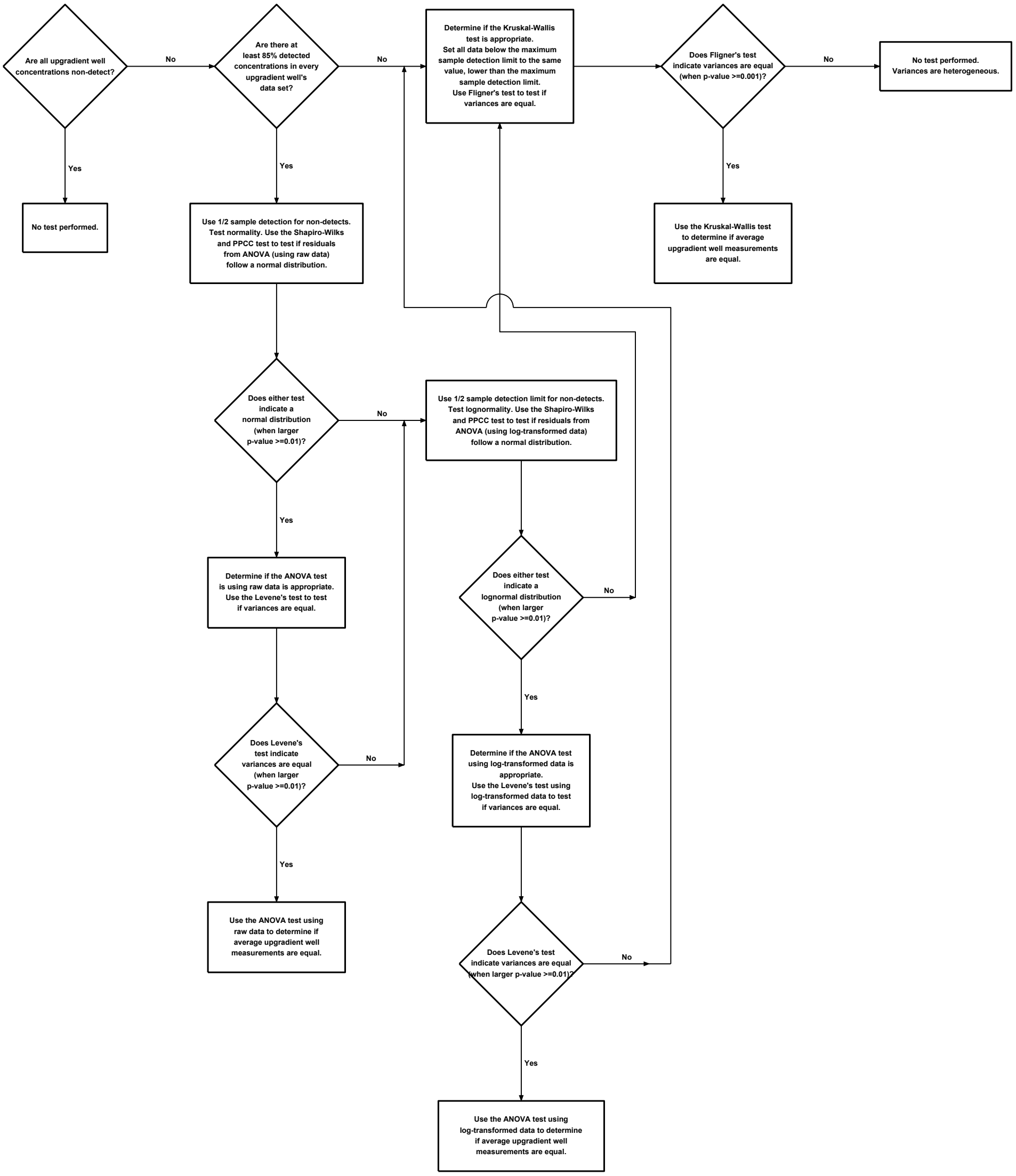
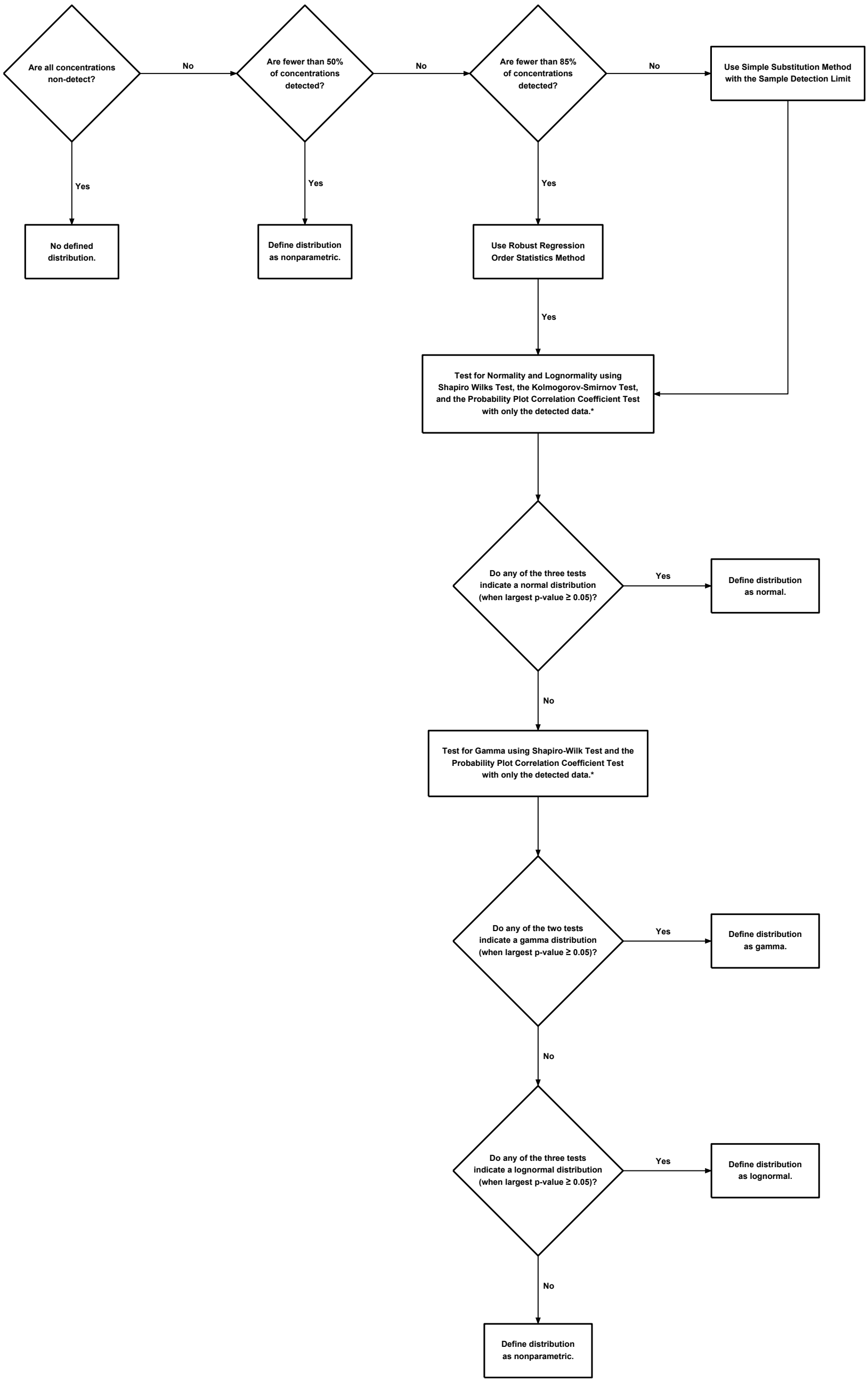


Figure 1

PROCESS FOR SELECTING ANOVA OR KRUSKAL-WALLIS TEST TO COMPARE UPGRADIENT WELL AVERAGES

PROJECT: 5164E	BY: AJD	REVISIONS
DATE: JAN 2019	CHECKED: PJB	

GOLDER ASSOCIATES USA, INC.



Note:

* - Distributional tests can not be performed for the following cases:

1. For a data group with fewer than five detected samples, the Kolmogorov-Smirnov Test and the Probability Plot Correlation Coefficient Test can not be performed using only the detected concentrations.
2. For a data group with fewer than four detected samples, the Shapiro-Wilks Test can not be performed using only the detected concentrations.

Figure 2

PROCESS FOR DEFINING A DISTRIBUTION FOR A DATA SET

PROJECT: 5164E	BY: AJD	REVISIONS
DATE: JAN 2019	CHECKED: PJB	

GOLDER ASSOCIATES USA, INC.