REPORT

Coal Combustion Residuals Groundwater Statistical Method Certification, Revision 2

Great River Energy - Coal Creek Station

Submitted to:

Great River Energy

Coal Creek Station

2875 Third Street SW, Underwood, North Dakota

Submitted by:

Golder Associates USA Inc.

7245 W Alaska Drive, Suite 200, Lakewood, Colorado, USA 80226

+1 303 980-0540

21451024C-10-R-0

November 29, 2021

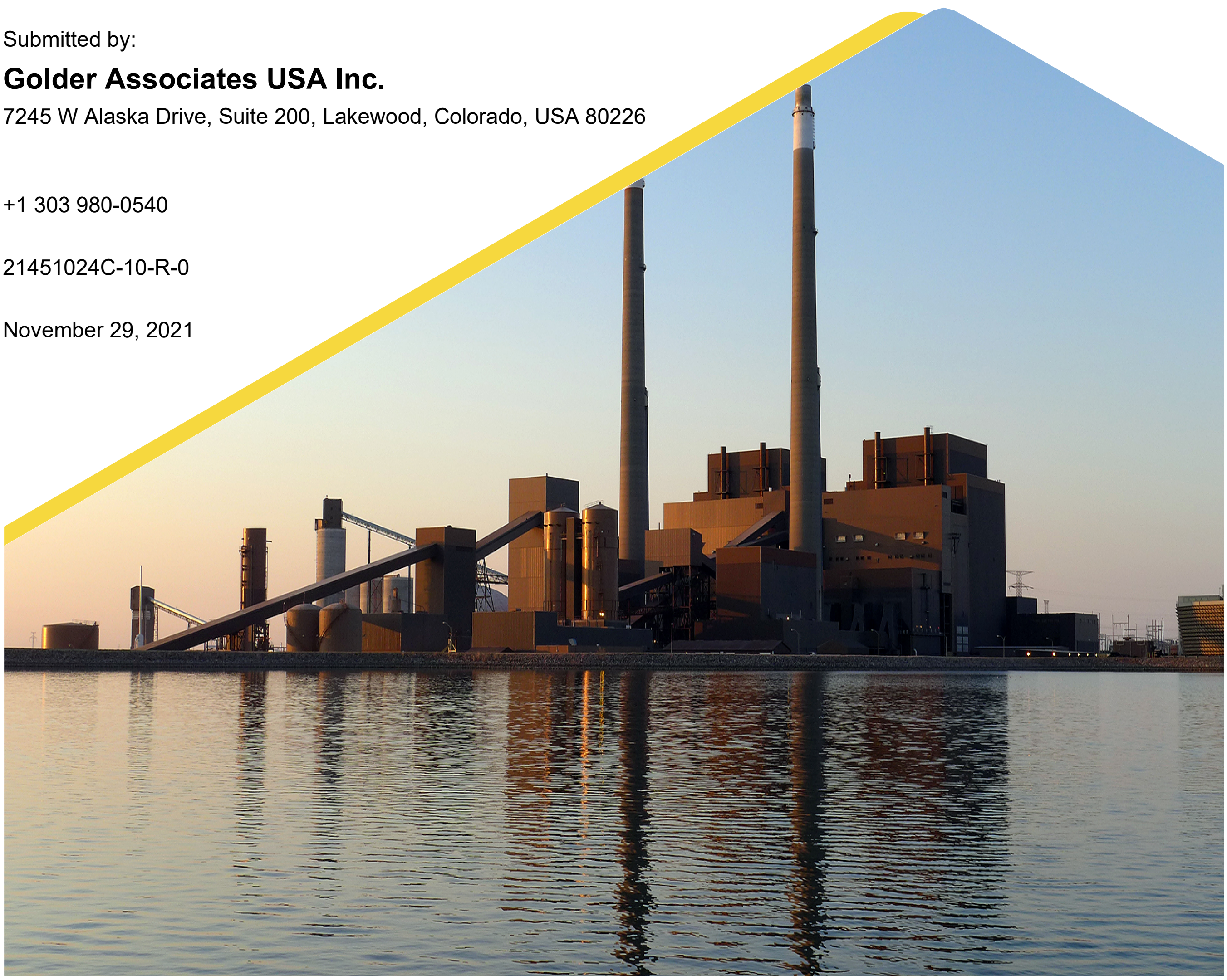

Table of Contents

TABLES

1.0 INTRODUCTION

Golder Associates USA Inc. (Golder), a member of WSP, prepared this revision to the Coal Combustion Residuals (CCR) Groundwater Statistical Method Certification Report on behalf of Great River Energy (GRE) to certify that the statistical methodology to be employed at GRE's Coal Creek Station (CCS) is appropriate and meets the requirements of the United States Environmental Protection Agency's (USEPA's) CCR Rule, 40 Code of Federal Regulations (CFR) Part 257.93 (USEPA 2015).

1.1 Purpose

The CCR Rule establishes specific requirements for statistical analysis of groundwater monitoring networks in 40 CFR 257.93 (Groundwater Sampling and Analysis Requirements). Per 40 CFR 257.93(f), owners or operators of CCR units must select a statistical method to evaluate groundwater monitoring data for the constituents specified in Appendix III and Appendix IV of the CCR Rule. This report serves as the documentation of the required certification of the selected statistical methodology in accordance with 40 CFR 257.93(f)(6).

This document may be periodically revised based on a review of changes to guidance documents provided by the USEPA or other governing entities, changes to the facility groundwater monitoring plan, as well as revisions to the language of the rule itself. A revision history is provided in Section 5.0.

1.2 Site Background

CCS is located in McLean County, approximately 10 miles northwest of Washburn, North Dakota. CCS has four CCR facilities regulated under the CCR Rule. The CCR units include the following:

- Drains Pond System CCR Surface Impoundment (Drains Pond System)
- Upstream Raise 91 CCR Surface Impoundment (Upstream Raise 91)
- Upstream Raise 92 CCR Surface Impoundment (Upstream Raise 92)
- Southeast Section 16 CCR Landfill (Southeast 16)

1.3 Constituents

Baseline samples (a minimum of eight) are collected for each well within the system and analyzed for both Appendix III and Appendix IV constituent lists associated with the CCR Rule (Table 1). Additionally, the following field parameters are recorded: pH, specific conductivity, turbidity, and temperature. Following collection of baseline samples, semi-annual detection monitoring samples are analyzed for the Appendix III constituent list and field parameters. If assessment monitoring is required, samples will be analyzed for the Appendix IV constituent list in accordance with CCR Rule requirements. During the baseline period and all subsequent samples collected as part of the CCR sampling program, metals will be analyzed as total recoverable metals (i.e., samples will not be filtered).

The statistical analyses described in the following sections will be performed for each of the Appendix III constituents listed in Table 1 during detection monitoring, and for the Appendix IV constituents in the event of assessment monitoring. Apart from field pH, statistical analysis will not be conducted on field constituents (specific conductivity, turbidity, and temperature). There is a high potential for false positive results with field measurements due to sample collection inconsistency (e.g., changes in field conditions and sampling personnel), equipment calibration variability, and other causes of measurement variability.

Table 1: CCR Constituent List

Appendix III Constituents
Boron
Calcium
Chloride
Fluoride
pH (Field)
Sulfate
Total Dissolved Solids (TDS)

Appendix IV Constituents
Antimony
Arsenic
Barium
Beryllium
Cadmium
Chromium
Cobalt
Fluoride
Lead
Lithium
Mercury
Molybdenum
Radium-226 and -228 Combined
Selenium
Thallium

2.0 STATISTICAL METHODOLOGY

The purpose of groundwater monitoring in conjunction with the CCR Rule is to determine if the regulated CCR units are impacting groundwater. To determine if an impact has occurred, recent groundwater data from each well will undergo statistical analysis. The statistical methodology outlined within this document was selected in accordance with 40 CFR 257.93 of the CCR Rule and the *Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities, Unified Guidance* (Unified Guidance, USEPA 2009). Additional consideration was given to recommendations prepared by the Electric Power Research Institute (EPRI) (2015).

Baseline diagnostic tests will be conducted using baseline samples collected at each well. Statistical analysis will be conducted using Sanitas Technologies (Sanitas) (2014) or similar statistical analysis software.

During detection monitoring, comparative statistical analysis will be performed after each monitoring event to identify whether concentrations are potentially statistically significant in comparison to baseline values. The results will be reported annually. If assessment monitoring is conducted, comparative statistical analysis will be conducted following the collection of the specified number of samples.

The baseline data for all constituents, as determined by the monitoring program, will be reviewed periodically (approximately every two to three years based on a semi-annual sampling frequency during detection monitoring) to determine if recent results that are not statistically significant in comparison to baseline values can be incorporated into an updated baseline period (specifics of the baseline update procedure are discussed in depth in Section 2.2.6).

2.1 Intrawell Methodology

Intrawell statistics (comparing compliance data to baseline data within a single well) were identified as an appropriate method for groundwater monitoring at CCS. This method is based on the following factors outlined in the Unified Guidance (USEPA 2009). The Unified Guidance provides procedures for both interwell and intrawell methods: “Groundwater detection monitoring involves either a comparison between different monitoring stations (i.e., downgradient compliance wells versus upgradient wells) or a contrast between past and present data within a given station (i.e., intrawell comparisons)” (USEPA 2009). The Unified Guidance further identifies circumstances in which intrawell comparison may be preferred over interwell comparison. Two situations where intrawell statistics may be preferable to interwell statistics are where spatial variability is present, and where legacy impacts not stemming from current facilities or handling practices are known to have occurred. Across the monitoring units, CCS has both spatial variability and historical impacts from legacy units.

2.1.1 Spatial Variability

The Unified Guidance indicates that where spatial variability is evident, “the statistical approach would need to be modified so that distinct wells are treated as individual populations with statistical testing being conducted separately at each one (i.e., intrawell comparisons)” (USEPA 2009). Per the Unified Guidance, “Confounding results stemming from spatial variability can be eliminated” through the use of intrawell statistics (USEPA 2009). Specifically, the Unified Guidance notes that “substantial differences in mean concentration levels can *invalidate* interwell, upgradient-to-downgradient comparisons and point instead toward *intrawell* tests” (USEPA 2009, original emphasis included).

Due to the heterogeneity of the glacial geology at CCS, spatial variability is displayed throughout the groundwater monitoring networks. Use of intrawell statistics allows comparisons to be made on a per well basis, which avoids comparisons being made between differing geologic/geochemical conditions. The Unified Guidance identifies spatial variability as “statistically identifiable differences in mean and/or variance levels (but usually means) across the well field (i.e., spatial non-stationarity) (USEPA 2009).” Upgradient conditions were compared using statistical processes to determine spatial variability.

Three statistical lines of evidence demonstrated the geochemical spatial variability across the site. Side-by-side box plots of the Appendix III parameters for the upgradient wells at each facility were reviewed. The Unified Guidance notes that side-by-side box plots “can be further employed to screen for possible spatial variation in mean levels” (USEPA 2009). Further, “the extent to which apparent differences in mean levels seem to be real rather than chance fluctuations can be examined” through the comparison of side-by-side box plots (USEPA 2009). The side-by-side box plots display variable water quality at each facility.

Analysis of variance (ANOVA) testing provides further confirmation of upgradient spatial variability, as described in the Unified Guidance (USEPA 2009). Determination of the distribution for ANOVA testing is based on the residuals, or the calculated difference between the observed values and the values predicted by the ANOVA model. ANOVA testing was conducted for data collected from the upgradient locations. For each of the seven Appendix III parameters, either variation (in the case of parametric ANOVA tests for parameters with residuals exhibiting normal or transform-normal distributions) or difference (for non-parametric ANOVA tests for parameters with residuals exhibiting non-normal distributions) is indicated at the 5% significance levels (α = 95%). The Appendix III parameters exhibit ANOVA significance between upgradient wells, indicating statistically significant geochemical spatial variability. As noted by the Unified Guidance, “substantial differences among the mean levels at a set of uncontaminated sampling locations [such as the tested upgradient wells] are suggestive of natural spatial variability" (USEPA 2009). ANOVA evaluation of the upgradient wells identified significant difference between well means or medians. The Unified Guidance indicates that pooling should not occur among well means or medians determined to be significantly different using ANOVA (USEPA 2009) and therefore interwell statistical analyses are not recommended in this situation.

A major ion trilinear diagram (Piper diagram) developed from water chemistry results collected from the upgradient wells in the CCR groundwater program for each of the four CCR monitoring units (Upstream Raise 91, Upstream Raise 92, Southeast 16, and the Drains Pond System), shows the variability in water types across the site in upgradient wells.

As recommended by the Unified Guidance, the evidence of spatial variability amongst the upgradient wells has driven the selection of an intrawell statistical approach for each of the site facilities.

2.1.2 Historical Impacts

In addition to the identification and evaluation of spatial variability discussed above, the Unified Guidance indicates that intrawell testing is also useful when there are historical impacts. "Another advantage using intrawell background is that a *reasonable baseline* for tests of future observations can be established at historically contaminated wells. In this case, the intrawell background can be used to track the onset of even more extensive contamination in the future" (USEPA 2009, original emphasis included).

Past CCR depositional facilities are known to have impacted the subsurface hydrology. Those facilities have been removed and new facilities with engineered lining systems constructed. The use of intrawell statistics allows future potential changes in water chemistry to be more readily tracked in relation to current CCR facilities at CCS.

2.2 Methodology for Baseline Diagnostic Tests

2.2.1 Initial Data Review and Non-detect Handling

Initially, data will be plotted on time-series graphs to assess the temporal variability of the data and to visually screen for potential outliers. Temporal variability can be caused by seasonality, changes to the monitored system, changes to the analytical method, recalibration of instruments, and anomalies in the sampling method (USEPA 2009).

Non-detect (ND) values are results where the constituent is not detected at a concentration above the Practical Quantitation Limit (PQL). The PQL is the lowest concentration that can be reliably measured within the specified limits of precision and accuracy during routine laboratory operating conditions. ND values will be managed following the approach recommended by the Unified Guidance (USEPA 2009):

- For sample populations with less than 15% ND, direct substitution of the NDs with ½ the PQL.
- For sample populations with greater than 15% but less than or equal to 50% ND, the Kaplan-Meier approach will be used to estimate the mean and standard deviation of the population.
- For sample populations with greater than 50% ND, non-parametric approaches will be used to assess the data, with direct substitution of the NDs with the PQL.

2.2.2 Data Distribution

Parametric statistical tests are based on the assumption that the data are normally-distributed or can be transformed to a normal distribution. The distribution of the data will be tested for normality using the Shapiro-Wilk normality test with at 95% confidence level (or the Shapiro-Francia test when there are more than 50 results within the dataset). Each constituent from each well will be analyzed separately. Datasets found to be non-normal will be tested for other distributions using the “Ladder of Powers” described by the Unified Guidance and transformed accordingly. Following transformation, parametric statistical methods will be performed on the normally or transform-normally distributed data. Non-parametric statistics will be used for datasets that do not show normal or transform-normal distributions, or for populations with greater than 50% ND.

2.2.3 Outlier Analysis

In accordance with the Unified Guidance, data points will be identified as outliers if the value was an “extreme, unusual-looking measurement’ and ‘inconsistent with the distribution of the remaining measurements.” The Unified Guidance recommends testing for outliers within baseline data, but cautions against removal of outliers, unless a likely error or specific discrepancy can be identified, such as recordkeeping errors, unusual procedures or conditions during sampling or at the laboratory, inconsistent sample turbidity, or values significantly outside the range of other results. In accordance with the Unified Guidance, apparent outliers will be periodically revisited even if initially removed due to the propensity of groundwater chemistry to change over time.

Outliers will be evaluated and identified through visual inspection and the USEPA-recommended Dixon’s Test for statistical outliers (or Rosner’s Test when there are greater than 25 samples). Dixon’s Test and Rosner’s Test assume that all data values, except the suspected outlier, are normally distributed or can be transformed to fit a normal distribution. Consequently, visual inspection of concentrations over time is important in screening for outliers. The effect of removing outliers from the baseline data will usually be to lower statistical limit (SL) (due to a reduction in the standard deviation of the dataset), resulting in a more conservative SL and improving the odds of detecting increasing concentration levels.

Outliers will be managed as follows:

- Any suspected outliers identified through statistical analysis or visual methods will be reviewed (i.e., through evaluation of the associated analytical report, laboratory narrative, associated laboratory quality assurance/quality control information, and/or field notes) before removal from the dataset to determine if any systematic or systemic errors were responsible for the noted anomalous readings. Rejected data points will not be included in the baseline dataset.
- The rationale for the removal of any outliers will be documented. The majority of outliers will likely be isolated values that can be attributed to inconsistent sampling or analytical chemistry methodology resulting in laboratory contamination or other anomalies.
- If an outlier is removed, the normality test (Section 2.2.2) will be re-run to determine if the dataset is normally-distributed or transform-normal without the outlier.

2.2.4 Trend Analysis

Most statistical tests assume concentrations do not demonstrate temporal correlation. The Sen’s Slope methodology is an intrawell statistical analysis of increases or decreases in measured concentrations over time measured by calculating the slope of the linear relationship of concentration levels and time. Sen’s Slope Methodology is paired with the Mann-Kendall test to determine the statistical significance of the calculated Sen’s Slope. The methodology involves examining all possible pairs of measurements in the dataset and scoring each pair to determine if a trend exists. The test will be conducted using a target confidence level of 99%.

If temporal trends are identified within the dataset, the data will be adjusted to account for the trends, the time period used for the baseline statistics will be reassessed, and/or an alternative statistical method will be used to establish a limit. In the case of downward trends, no adjustments may be considered appropriate and a particular constituent may not be considered for statistical analysis until further data is collected.

2.2.5 Seasonality

Seasonal temporal variability can mask changes in groundwater chemistry. Time-series plots will be observed for visual signs of seasonality and, once enough data has been collected at each well, the data will be evaluated for seasonal variations using the Kruskal-Wallis test. The Kruskal-Wallis test is an intrawell evaluation performed individually for each constituent at each well and requires at least three occurrences of each prospective season with results to calculate the statistic. Datasets are analyzed for seasonality with two seasons assumed per calendar year based on site observations, with seasonal start dates set on April 1 and October 1. Datasets that are found to have seasonality will be deseasonalized in accordance with the method described in the Unified Guidance for subsequent statistical analysis.

2.2.6 Updating the Baseline Period

The Unified Guidance recommends updating the baseline period every two to three years when sampling frequency is semi-annual, or every four to eight collected samples. A baseline update will include a review of any revisions to federal and state regulations and USEPA statistical guidance documents that may have been promulgated since the previous baseline statistical analysis. The baseline period for a specific constituent will not be considered for an update if verified statistically significant increases (SSIs) have been identified for that constituent unless a successful alternative source demonstration has been made.

Prior to inclusions of more recent data in an updated baseline period, a Wilcoxon Rank-Sum test will be conducted. The Wilcoxon Rank-Sum test, also referred to as the Mann-Whitney test, determines if measurements from one population are statistically significantly higher or lower than another population. The test is non-parametric, namely the data being tested are not assumed to fit a specific distribution, such as a normal distribution. When the baseline period is updated in the future, the Wilcoxon Rank-Sum test will be used to compare data from the current baseline period with the more recent data that are intended to be reclassified and included in the updated baseline period. The test will be conducted at a 95 to 99 percent confidence level. If the two datasets are drawn from the same population, then the results of the test support updating the prior baseline data set with the recent data. After the new data are incorporated into the dataset, the baseline diagnostic tests outlined in Section 2.2 will be conducted.

If the Wilcoxon Rank-Sum test detects a significant difference between two sample populations, additional data review will be necessary. The data will be reviewed to determine whether a gradual trend or other change not stemming from a release from the facility has occurred that was not detected during comparative statistical analysis. At the time of the baseline update, some earlier baseline data might need to be removed from the updated baseline period. Removal of earlier baseline data ensures that future statistical analysis is based on current groundwater conditions, rather than on outdated measurements of groundwater chemistry. Additionally, in cases where an SSI was identified and a successful alternative source demonstration was conducted, exclusion of some data points might be appropriate in order to update the baseline dataset. Alternatively, outliers identified in the previous baseline period (as described in Section 2.2.3) will be re-incorporated into the dataset and reevaluated as potential outliers during the baseline update unless the outlier(s) were removed due to sampling, laboratory, or other determinant error.

3.0 DETECTION MONITORING

3.1 Statistical Limits

Either a parametric or non-parametric method will be used to generate the baseline SL for each constituent. The statistical method may vary between constituents and will be selected based on the percent of ND values in the baseline period and the baseline data distribution for each constituent at each well, in accordance with the Unified Guidance (USEPA 2009).

For those constituent-well pairs where concentrations of a given analyte are normally or transform-normally distributed and have equal to or greater than 50% detections, either intrawell parametric prediction limits (PLs) or Shewhart-CUSUM (cumulative summation) control charts will be used.

For detection monitoring, intrawell parametric PLs compare individual compliance measurements to a baseline limit. The general equation for an intrawell parametric PL based on normal or transform normal populations is given as:

$$PL = \bar{x}_{BG} + \kappa s_{BG}$$

Where the PL is calculated using the mean ($\bar{x}$) and standard deviation (s) of the baseline dataset, and κ is a multiplier based on the number of compliance observations as part of the retesting scheme. If used, PLs would be calculated using a 1-of-2 retesting scheme.

The Unified Guidance notes that Shewhart-CUSUM control charts use two separate evaluation procedures. The Shewhart portion is similar to a parametric PL, comparing compliance measurements to a baseline limit. The CUSUM portion of the test analyzes new measurements against prior compliance measurements. The mean ($\bar{x}$) and standard deviation (s) of the baseline dataset are used to calculate the SL, by the following equation:

$$SL = \bar{x}_{BG} + h s_{BG}$$

Per the Unified Guidance, h is the standardized control limit.

Where the concentrations of a given constituent-well pair are not normally or transform-normally distributed, or contain less than or equal to 50% detections, a non-parametric PL limit will be used. The non-parametric limit will be assigned as the highest detected value (excluding outliers) or the highest PQL, whichever is greater.

In the case of increasing trends within the baseline period, the data will be adjusted to account for the trends where a source other than the facility can be identified and/or an alternative statistical method will be used to establish a limit. In the case of downward trends, no adjustments may be considered appropriate, and a constituent-well pair may not be considered for statistical analysis until further data is collected; therefore, the trend test will serve as the alternative method for constituents with decreasing trend until the trend stabilizes.

3.2 Comparative Statistical Analysis

Comparative statistical analysis will be conducted following each detection monitoring event. For both Shewhart-CUSUM limits and non-parametric PLs, the comparative statistical analysis will consist of a comparison of detection monitoring results (the recent analytical results for each monitoring event performed after the baseline data period) to the SL calculated from the baseline data.

The following definitions will be used in discussion of the comparative statistical analysis:

- Elevated CUSUM – occurs when the calculated CUSUM value is greater than the Shewhart-CUSUM limit established by the baseline statistical analysis, but the analytical result does not exceed the Shewhart-CUSUM limit. An elevated CUSUM is an indication that concentrations are gradually increasing and that analytical results may exceed the Shewhart-CUSUM limit in the future. For elevated CUSUMs in the case of two-tailed analysis for field-measured pH, the CUSUM value may also be below the lower Shewhart-CUSUM limit established by the baseline statistical analysis
- Potential Exceedance – is defined as an initial elevated CUSUM or an initial analytical result that exceeds the parametric PL, the Shewhart-CUSUM limit, or the non-parametric SL established by the baseline statistical analysis. Confirmatory resampling will determine if the potential exceedance is a false-positive or a verified statistically significant increase (SSI). Non-detect results that exceed either the parametric PL, the Shewhart-CUSUM limit, or the non-parametric SL are not considered potential exceedances.
- False-positive – is defined as an analytical result that exceeds the SL that can clearly be attributed to laboratory error, changes in analytical precision, or is invalidated through confirmatory re-sampling. False-positives are not used in calculation of any subsequent CUSUMs.
- Confirmatory re-sampling – is designated as the next scheduled sampling event.
- Verified SSI – is interpreted as two consecutive exceedances (the original sample and the confirmatory re-sample for analytical results, two consecutive elevated CUSUMs, or a combination of an analytical result above the SL and an elevated CUSUM in either event order) for the same constituent at the same well.

The detection monitoring program has been developed to identify potential SSIs over baseline values for the Appendix III constituents. This determination will be made within 90 days of receiving the finalized laboratory analytical report(s) for each sampling event and completion of data quality review as necessary to address questions concerning the validity of sampling methods or laboratory analyses. A potential exceedance will not be considered a verified SSI until confirmatory re-sampling is performed. Confirmatory re-sampling will occur during the next regularly scheduled sampling event.

3.3 Alternative Source Demonstrations and Initiating Assessment Monitoring

If a verified SSI is identified in a downgradient well for an Appendix III constituent as part of the detection monitoring program, GRE will establish an assessment monitoring program meeting the requirements of 40 CFR 257.95 of the CCR Rule. Alternatively, GRE may demonstrate that a source other than the regulated CCR facilities caused the verified SSI, or that the verified SSI was a result of an error in sampling, analysis, statistical evaluation, or natural variability in the groundwater quality that failed to be captured during baseline data collection (40 CFR 257.94(e)(2)). A report documenting the alternative source demonstration (ASD) will be certified by a Professional Engineer registered in the state of North Dakota. A copy of the report will be placed in the Site's Operating Record within 90 days of the SSI determination. If a successful demonstration is made and documented, GRE will continue with the detection monitoring program. If, after 90 days, a successful ASD is not made, GRE will initiate an assessment monitoring program as described in Section 4.0.

4.0 ASSESSMENT MONITORING

An assessment monitoring program will be initiated in the event of a verified SSI of an Appendix III constituent in a downgradient well, unless a successful demonstration is made that an alternative source affected the groundwater chemistry, or that the verified SSI was a result of an error in sampling, analysis, statistical evaluation, or natural variability in the groundwater quality that was not captured during the baseline data collection period. Assessment monitoring samples will be analyzed for the Appendix IV constituent list in accordance with CCR Rule requirements, including at least one annual sample for the complete Appendix IV constituent list and at least two semi-annual samples for detected Appendix IV parameters based on the annual sample results. At a minimum, three samples will be collected for detected Appendix IV parameters with one sample collected for the complete Appendix IV parameter list annually.

The statistical comparisons used in assessment monitoring require baseline diagnostic tests to be completed prior to comparative statistics. Baseline diagnostic tests for assessment monitoring mirror those for detection monitoring and are discussed in further detail in Section 2.2, Section 4.2.1, and Section 4.2.2.

4.1 Establishment of Groundwater Protection Standards

The assessment monitoring program will comply with 40 CFR 257.95. Within 90 days of triggering the assessment monitoring program, GRE will sample and analyze each well in the affected monitoring network for the Appendix IV constituents listed in Table 1. Groundwater protection standards (GWPS) will be established for each detected Appendix IV constituent. Per the CCR Rule (40 CFR 257.95(h)), the GWPS must fall within one of the following categories:

- For constituents for which a maximum contaminant level (MCL) has been established by the USEPA (40 CFR 141.62 and 141.66), the MCL for that constituent will be the GWPS, per 40 CFR 257.95(h)(1).
- For the following constituents, the following alternative limits apply, per 40 CFR 257.95(h)(2):
  - Cobalt – 0.006 milligrams per liter (mg/L)
  - Lead – 0.015 mg/L
  - Lithium – 0.04 mg/L
  - Molybdenum – 0.1 mg/L
- For constituents where the upgradient baseline concentration is higher than the levels identified in 40 CFR 257.95(h)(1) and 40 CFR 257.95(h)(2), an SL determined from the baseline data (derivation described below) will be the GWPS per 40 CFR 257.95(h)(3).

For those constituents where the upgradient baseline concentration is higher than the MCL or levels designated in 40 CFR 257.95(h)(2), the GWPS will be determined through statistical methods. The Unified Guidance provides two acceptable approaches for conducting statistical evaluations using non-MCL or specified limit based GWPS: a tolerance interval approach or a prediction interval approach. The Unified Guidance recommends using a tolerance interval approach over a prediction interval approach if the background (or historical) dataset is normally distributed, transform-normally distributed, or has a sufficient number of background (or historical) observations for a non-parametric tolerance interval. For this program, a tolerance interval approach will be the primary method used for determining the site-specific GWPS levels.

Per the Unified Guidance, a tolerance interval is “a concentration range designed to contain a pre-specified proportion of the underlying population from which the statistical sample is drawn.” A tolerance limit is a one-sided tolerance interval. The Unified Guidance recommends “an upper tolerance limit (UTL) based on both the background sample size and sample variability” for the site-specific GWPS. Tolerance limits will be produced using data collected from multiple upgradient wells that can be “reasonably combined”, giving a pooled background data set (USEPA 2009). Tolerance limits are defined with two coefficients, representing the portion of the population that the interval is intended to cover (the coverage, γ) and the degree of confidence that the interval reaches the coverage (the confidence level, 1-α). A coverage of 95% (γ = 0.95) and a confidence of 95% (α = 0.05) will be used. For background datasets with a normal distribution, the formula for the UTL serving as the GWPS is as follows:

$$UTL = \bar{x} + \tau_{(n,0.95,0.95)} \cdot s$$

Within the equation, n is the number of background measurements, $\bar{x}$ is the background sample mean, s is the background sample deviation, and τ is a tolerance factor based on the number of background measurements, n, with a coverage of 95% and confidence of 95%. Tabulated τ values are available in Table 17-3 in Appendix D of the Unified Guidance (USEPA 2009).

For background datasets with non-normal distributions, particularly for cases with a significant portion of non-detect observations, a non-parametric tolerance interval approach will be used. The UTL for the non-parametric tolerance limit will be set at the highest detected value (excluding outliers) or the highest PQL, whichever is greater.

If tolerance intervals cannot be used to calculate the GWPS (for example, small non-normal datasets or a dataset with a high percentage of non-detects), then a prediction interval (or PL) method will be used. For assessment monitoring, the Unified Guidance suggests using a prediction interval for a future mean for normal or transform normal datasets and a prediction interval for a future median for data with a high percent of NDs or non-normally distributed data.

A facility-wide GWPS based on the pooling of upgradient well data is preferred; however, the Unified Guidance recognizes that an intrawell approach resulting in the development of well-specific GWPS may be required if there are individual well spatial differences that need to be considered. If data review indicates an intrawell approach is more appropriate, well-specific tolerance or PLs will be developed based on the historical data for a given well.

4.2 Comparative Statistics

To determine if Appendix IV constituents have statistically exceeded the associated GWPS, the approaches discussed in the following sections will be used.

4.2.1 Maximum Contaminant Levels and Specified Limits Based GWPS

For those constituents with upgradient baseline concentrations below the maximum contaminant levels (MCLs) or pre-defined GWPS limits specified in 40 CFR 257.95(h)(2), a confidence interval approach will be used for determining compliance. Per recommendations provided by the Unified Guidance and detailed by EPRI (2015), a confidence interval statistically defines the upper and lower bound (the upper and lower confidence limit) of the true mean associated with a groundwater population. The Unified Guidance recommends confidence intervals for assessment monitoring. A confidence interval statistically defines the confidence range and the upper and lower bound of a true mean, median, or other statistical measure of a compliance monitoring dataset. Confidence intervals identify statistically significant levels (SSLs) through comparison against a fixed standard, namely the GWPS defined as the MCL or the limits specified in 40 CFR 257.95(h)(2).

The specific type of confidence interval will be based on three factors:

1) the detection frequency of compliance data

2) the distribution of the compliance data

3) if the compliance data displays a statistically significant trend

Prior to constructing confidence intervals, initial data review will be conducted on the compliance data for determining the percentage of ND (i.e., the detection frequency). Substitution of ND will be handled using the recommendations provided by the Unified Guidance, based on the percentage of ND within the dataset, as described below:

- For sample populations with less than 15% ND, direct substitution of the NDs with ½ the PQL will be used for estimating the mean and standard deviation of the population for construction of the parametric confidence interval if the dataset is normal or transform normal.
- For sample populations with greater than 15% but less than or equal to 50% ND, the Kaplan-Meier approach will be used to estimate the mean and standard deviation of the population for construction of the parametric confidence interval if the dataset is normal or transform normal.
- For sample populations with greater than 50% ND, non-parametric confidence intervals on the median will be used to assess the data, with direct substitution of the NDs with the PQL.

Following initial data review of the compliance data, the distribution of the data will be tested for normality using the Shapiro-Wilk normality test with a 95% confidence level. Each constituent from each well will be analyzed separately. Datasets found to be non-normal will be tested for other distributions using the 'Ladder of Powers' described by the Unified Guidance and transformed accordingly. Following transformation, parametric statistical methods will be performed on the normally or transform-normally distributed data. Non-parametric statistics will be used for datasets that do not show normal or transform-normal distributions, or for populations with greater than 50% ND.

A parametric confidence interval around the mean will be used if the compliance dataset has less than 50% ND and is normally distributed or transform-normally distributed. A confidence (α) of 95% will be used for calculating the Upper Confidence Limit (UCL) and Lower Confidence Limit (LCL) of the compliance data. To calculate a confidence interval around the mean, a minimum of four independent compliance samples are required.

A non-parametric confidence interval around the median will be used for datasets that do not show normal or transform-normal distributions, or for populations with greater than 50% ND. To calculate a non-parametric confidence interval around the median, a minimum of three independent compliance samples are required. However, additional samples will increase the confidence (α) of the tests.

For determination of SSLs, the LCL is of primary interest. A confidence interval will only be considered statistically above the associated GWPS if both the UCL and LCL exceed the GWPS. If only the UCL exceeds the GWPS while the LCL remains below the GWPS, the results of the test are considered inconclusive. The Unified Guidance recommends results of this category to be interpreted as "in compliance" and not consider the results to represent an SSL. If both the UCL and LCL are below the GWPS, the data are considered not statistically significant.

Confidence intervals will only be constructed on data that do not display statistically significant trends. Compliance data will be tested for trends using the Sen's Slope methodology (as described in Section 2.2.4) prior to constructing confidence intervals. If the compliance dataset is found to display a statistically significant trend, a confidence band will be constructed around the estimated trend line. Per the Unified Guidance, a confidence band is “essentially a continuous series of confidence intervals estimated along every point of the trend.”

4.2.2 Background-Based GWPS

If a tolerance interval approach is used for a GWPS, a confidence interval approach will be used for determining compliance, similar to the approach for constituents with GWPS set at the MCL or pre-defined GWPS limits. Confidence intervals will be compared to the upper tolerance interval to identify SSLs. Confidence intervals will be calculated following the steps detailed in Section 4.2.1. To calculate a confidence interval, at least four independent compliance samples are required. A confidence interval will only be considered statistically above the associated GWPS if both the upper and lower confidence limits exceed the GWPS defined by the tolerance interval.

If a prediction interval approach is used for a GWPS, compliance data will not be assessed with confidence intervals. For comparisons to a parametric prediction interval for future means, a minimum of four compliance measurements are averaged to compare to the GWPS, with statistical significance occurring when the calculated comparative mean is greater than the GWPS. For comparisons to non-parametric prediction intervals for future medians, the median of at least three measurements will be calculated and compared to the GWPS, with statistical significance occurring when the calculated comparative median is greater than the GWPS.

4.3 Returning to Detection Monitoring, Assessment Monitoring ASDs, and Assessment of Corrective Measurements

If the concentrations of constituents listed in Appendix III and Appendix IV of the rule are shown to be at or below baseline SLs derived from detection monitoring statistical approaches for two consecutive sampling events, GRE will return to detection monitoring. This determination of baseline values is based on the statistical evaluation procedure established for detection monitoring (Section 3.0).

If the concentrations of any of the Appendix III and Appendix IV constituents are above baseline SLs, but the Appendix IV constituents are below the established GWPS, GRE will continue with assessment monitoring. If one or more Appendix IV constituents are detected at statistically SSLs above the established GWPS, GRE will place a notification of the exceedance in the Site Operating Record and follow the course of action outlined in 40 CFR 257.95(g) of the CCR Rule.

As an alternative, GRE may demonstrate that a source other than the regulated CCR facilities caused the SSL, or that the SSL was a result of an error in sampling, analysis, statistical evaluation, or natural variability in the groundwater quality that failed to be captured during baseline data collection (40 CFR 257.95(g)(3)(ii)). A report documenting the ASD will be certified by a Professional Engineer registered in the state of North Dakota. A copy of the report will be placed in the Site’s Operating Record within 90 days of the SSL determination. If a successful demonstration is made and documented, the owner or operator must continue monitoring in accordance with the assessment monitoring program pursuant to 40 CFR 257.95 and may return to detection monitoring if the constituents in Appendix III and Appendix IV are at or below background as described above. If, within 90 days, a successful ASD has not been made, GRE will initiate an assessment of corrective measures (ACM) within 90 days in accordance with 40 CFR 257.96. The ACM will be completed within 90 days of being initiated unless a demonstration is made for additional time to complete the assessment. This demonstration will be certified by a Professional Engineer registered in the state of North Dakota.

5.0 REVISION HISTORY

A history of revisions to this document:

Revision 0 – Published October 12, 2017.

Revision 1 – Published March 8, 2019. Revised to reflect:

1) Updates to the CCR Rule published in the Federal Register on July 30, 2018.
2) New CCR unit naming convention (Upstream Raise to Upstream Raise 92 and Ash Pond 91 to Upstream Raise 91).
3) Removal of calendar dates associated with the initial collection of background samples to not conflict with baseline sample collection dates for any future wells added to the program.

Revision 2 – Published November 17, 2021. Revised to reflect:

1) Correction to seasonality testing approach used for baseline establishment and baseline update.
2) Expansion of assessment monitoring statistical approach.
3) Minor grammatical and terminology corrections changed for increased clarity.

6.0 CERTIFICATION

Based upon the review described in this report, the undersigned certifies that the statistical methodology presented herein is appropriate for evaluating groundwater monitoring data associated with the regulated CCR facilities at CCS and meet the requirements of 40 CFR 257.93.

Golder Associates USA Inc.

REGISTERED PROFESSIONAL ENGINEER
TODD J. STONG
PE-6144
DATE
NORTH DAKOTA

Todd Stong, P.E.	Erin L. Hunder, PhD
Associate and Senior Consultant	*Senior Project Engineer*

TS/ELH/mb

Golder and the G logo are trademarks of Golder Associates Corporation

https://golderassociates.sharepoint.com/sites/140044/project files/6 deliverables/21451024c/reports/10-r-ccr_stats_cert_rev2/10-r-0/21451024c-10-r-0-ccr_stats_cert_rev2_29nov21.docx

7.0 REFERENCES

EPRI (Electric Power Research Institute). 2015. Groundwater Monitoring Guidance for the Coal Combustion Residuals Rule – 2015 Technical Report. November 2015.

Sanitas (Sanitas Technologies). 2014. Sanitas User Guide, Version 9.5.

USEPA (United States Environmental Protection Agency). 2009. *Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities Unified Guidance*, Office of Resource Conservation and Recovery, EPA-R-09-007, March 2009.

USEPA. 2015. Code of Federal Regulations Title 40 Part 257: Hazardous and Solid Waste Management System: Disposal of Coal Combustion Residuals from Electric Utilities. April 17, 2015.

USEPA. 2018. Code of Federal Regulations Title 40 Part 257: Hazardous and Solid Waste Management System: Disposal of Coal Combustion Residuals form Electric Utilities; Amendments to the National Minimum Criteria (Phase One, Part One). July 30, 2018.

GOLDER

MEMBER OF WSP

golder.com