REPORT

Statistical Analysis Plan

AEPCO Apache Generating Station
Cochise, Arizona

Submitted to:

Arizona Electric Power Cooperative, Inc.

3525 N. Highway 191, Cochise, Arizona USA 85606

Submitted by:

Golder Associates Inc.

4730 North Oracle Road, Suite 210, Tucson, Arizona, USA 85705

+1 520 888-8818

Registered Professional Engineer (Civil)
CERTIFICATE NO. 23335
KENT R. JOHNEJACK
DATE SIGNED 7/12/19
ARIZONA U.S.A.
Expires 06 / 30 / 2020

19116278

July 12, 2019

Distribution List

Apache Generating Station Operating Record

Arizona Electric Power Cooperative, Inc.

Golder Associates Inc.

Table of Contents

1.0 INTRODUCTION

This statistical analysis plan has been prepared in accordance with the coal combustion residuals (CCR) rules from United States Environmental Protection Agency (USEPA) that require the development of a statistical analysis plan to evaluate if CCR units are impacting downgradient groundwater quality. These rules, in whole or in part, require the statistical evaluation of the analytical data from sampling events to identify whether statistically significant increases (SSIs) in concentrations above background levels are present. Descriptions of acceptable statistical programs are provided in USEPA's document *Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities, Unified Guidance* (USEPA, 2009), which is commonly referred to as the "Unified Guidance."

A Statistical Analysis Plan (SAP) for a facility is based on review of the site-specific data from which an appropriate method is chosen for compliance with the CCR rules. Specifically, the SAP describes:

- The criteria used to justify the selected methodology
- A description of the statistical methodology that will be used
- The general statistical approach for detection monitoring based on data from groundwater monitoring wells
- If applicable, the approach for assessment monitoring and/or corrective action

This SAP is an update to a previous SAP (Golder, 2017) and provides a description of the statistical approach and methods used in accordance with the CCR rule reporting requirements [40 CFR §257.93(f)(6)]. Specifically, this SAP update provides details for statistical evaluation in the presence of a known upgradient source unrelated to the CCR Units.

This SAP has been prepared to evaluate groundwater quality data from Arizona Electric Power Cooperative's (AEPCO) Apache Generating Station (AGS). Two CCR units, designated as Ash Pond Unit 1 – 4 Multi-Unit and Scrubber Pond No. 2 Single Unit are in use at AGS.

2.0 STATISTICAL APPROACH FOR DETECTION MONITORING PROGRAMS

The statistical methodology meets the criteria referenced in the CCR Rule as well as the Unified Guidance.

2.1 Prediction Limits

Prediction limit tests are planned to be used for Appendix III constituents to represent anticipated future conditions. The prediction limit method is often preferred over control charts and tolerance limits. Relative to control charts, prediction limits are generally more conservative statistical limits and are a more straightforward procedure. Prediction limits are often preferred over tolerance limits, since tolerance limits are calculated to include a specified percentage of background observations (e.g., typically 99%), while prediction limits are calculated to include a specified percentage of future observations. As such, a tolerance limit is a backward-looking test procedure, which may be appropriate for statistically evaluating a population with a known set of constituents (e.g., part manufacturing). Conversely, a prediction limit is a forward-looking test procedure, which is appropriate for evaluating future events where the outcome is less certain, such as the case with groundwater monitoring.

2.1.1 Interwell Prediction Limits

Interwell prediction limits is a statistical method that evaluates groundwater quality from upgradient wells to construct statistical limits which are then used to evaluate downgradient well data. The Unified Guidance indicates that interwell evaluation is appropriate if:

- Upgradient well(s) are in the same aquifer as downgradient wells
- Sufficient groundwater velocity for flow from upgradient to downgradient wells in a reasonable time period is observed
- Background groundwater quality is expected to be statistically similar to downgradient groundwater (assuming no release)

Initially, newly regulated facilities implemented an accelerated background approach to collect a minimum of 8 background samples for statistical analysis, but the schedule for future sampling is based on semi-annual groundwater monitoring. AEPCOs background data collection is complete per the CCR Rule.

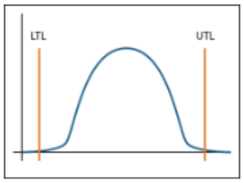
When it can be reasonably demonstrated that no contamination from current or historical practices at the regulated unit is present, the Unified Guidance recommends switching from interwell methods to intrawell methods.

2.1.2 Intrawell Prediction Limits

Intrawell Prediction Limits (PLs) compare the results from a particular well to statistical limits calculated from historical background groundwater quality measurements at the same well. In general, intrawell statistical methods are preferred over interwell methods, because they tend to produce more reasonable (i.e., more conservative) statistical limits and result in fewer false positive occurrences. Intrawell Prediction Limits are only appropriate if it can be documented that there has been no previous release. According to the Unified Guidance, one key advantage (for intrawell testing) is that confounding results due to spatial variability are eliminated, since all data used in an intrawell test are obtained from the same location. Intrawell comparisons should always be used when predisposal data are available (Gibbons, 2000).

In an ideal situation, upgradient and downgradient monitoring wells would be installed in a perfectly homogeneous aquifer that is chemically isotropic, where the only possible source of variability in groundwater quality between upgradient and downgradient wells would be the CCR unit which separates them. In reality, natural geochemical differences in groundwater quality exist between monitoring locations, which increases the risk that statistically significant differences reported between upgradient and downgradient locations are the result of spatial and/or hydrogeologic variability rather than CCR unit influence. In situations where the groundwater quality in the upgradient monitoring wells displays spatial variability, intrawell statistical limits are more conservative (i.e., intrawell statistical limits will reveal apparent statistical exceedances prior to interwell statistical limits, because the standard deviation for an intrawell statistical limit is based on the variability of a single well instead of a pooled set of data from several background wells). Intrawell statistical methods minimize the likelihood that spatial variability will contribute to invalid statistical limits, as long as there has been no previous release. Therefore, intrawell statistical evaluations are not recommended for facilities which are known to have impacts in downgradient wells.

2.1.3 Tolerance Limits



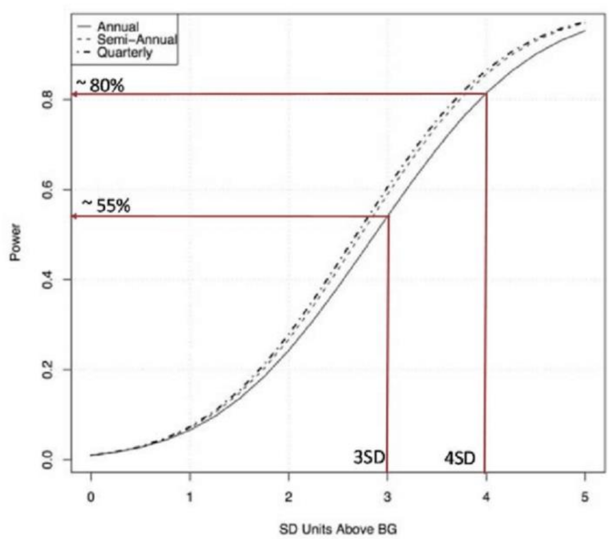
Tolerance intervals are statistical ranges typically constructed from on-site background data. Tolerance limits define the range of data that fall within a specified percentage with a specified level of confidence. The upper tolerance limit has been commonly used to establish a background threshold value; however, prediction limits are often favored for establishing a background threshold value in groundwater because they account for repeated measures. An upper tolerance limit (UTL) is designed to contain, but not exceed, a large fraction (that is, 95%, 99%) of the possible background concentrations, thus providing a reasonable upper limit on what is likely to be observed in background. Similarly, the lower tolerance limit (LTL) is designed to contain at most a certain percentage of the possible background concentrations, thus providing a reasonable lower limit on what is likely to be observed in background. The fraction to be contained or 'covered' by the limit is the coverage parameter and must be specified along with a desired confidence level. Tolerance limits explicitly account for the degree of variation in the background population and the size of the sample of measurements used to construct the limit. Tolerance limits and confidence limits are distinct, even though in some cases the one-sided upper limits for both methods are equivalent. It is important to note also that unlike prediction limits, tolerance limits are less flexible when incorporating formal retesting in detection monitoring. Assumptions to be considered before selecting tolerance limits for use in data evaluation include:

- Parametric tolerance limits assume the data follow a statistical distribution – typically normal (or can be normalized). If a transformation (for example, computing the logarithms of all of the data points is a transformation) is needed to normalize the measurements, the tolerance limit can be computed using the transformed values and then back transforming the results to get the final limit.
- Nonparametric tolerance limits do not assume normality or any particular distributional form (but generally require larger samples sizes than parametric tolerance limits).
- Tolerance limits assume the population is stable (or stationary) over the period of time during which measurements are collected. No obvious trends or temporal patterns should exist in the background data.
- Since tolerance limits typically involve interwell comparisons, their use in detection monitoring tests assumes minimal spatial variability.
- Tolerance limits assume that the measurements are independent.
- Comparison of compliance data against an upper tolerance limit assumes that the two populations being compared have similar variances. This condition can be assessed using a homogeneity of variance test but will be difficult to test directly unless you have at least four independent observations from each population (background and compliance).

2.2 Site-Wide False Positive Rates (SWFPR) and Statistical Power

The Unified Guidance recommends an annual site-wide false positive rate of 10%, which is distributed equally among the total number of sampling events. A site-wide false positive rate of 5% is targeted for each semi-annual sampling event. The USEPA also requires demonstration that the statistical methodology selected for a facility will provide adequate statistical power to detect a release, should one occur. The USEPA recommends statistical power of approximately 55% at 3 standard deviations or 80% at 4 standard deviations over the background mean (inset).

EPA Reference Power Curve



2.2.1 SWFPR

According to the Unified Guidance, the false positive rate (FPR) is defined as the probability that a test will falsely indicate impact even when no impact is present. False positives can be attributed to several causes, including natural variation in groundwater quality and/or variation in field or laboratory measurement and analysis processes. In general, the FPR increases in direct proportion to the number of comparisons being made, so a larger number of comparisons will increase the probability that impact will be indicated even though no impact has occurred. According to the Unified Guidance, an FPR of less than 5% is typically viewed as acceptable.

Several options are available for limiting the SWFPR of a facility where multiple comparisons are made: (1) limit the number of comparisons; (2) decrease the per test FPR; (3) allow retesting in cases where statistical exceedances are noted. With respect to Item 1, the CCR rules are prescriptive regarding the number of constituents that are statistically analyzed on a semi-annual basis, which means that reducing the number of comparisons is not possible. However, adjusting the per test FPR and allowing for retests are valid options for reducing false positives.

Decreasing the per test FPR will decrease the SWFPR. However, the USEPA mandates that the per test FPR must be at least 1% (i.e., minimum of a 99% confidence interval), in order to maintain the statistical power of the test used. Finally, the Unified Guidance also recommends the use of retesting strategies in cases where statistical exceedances are noted. Retesting (i.e., verification resampling) is a strategy that is employed in this SAP to reduce false positive observations.

2.2.2 Statistical Power

In addition to the SWFPR, the Unified Guidance also requires facilities to achieve adequate statistical power to detect a release. More specifically, USEPA recommends power of approximately 55% when concentration levels are 3 standard deviations above the background mean, or approximately 80% power at 4 standard deviations above the background mean.

The performance of a given testing strategy is displayed in Power Curves which are based on the particular statistical method chosen, the false positive rate associated with the statistical test, a resampling plan as well as the number of background samples available, and the monitoring network.

2.3 Background Data

Prior to utilizing background data for the purpose of constructing statistical limits, it is necessary to screen the data for anomalies such as high outliers (that can artificially elevate prediction limits) and for trends over time. Time series plots, box plots, and statistical outlier tests can be used to identify anomalous observations. Suspect observations are de-selected from prospective background data prior to establishing prediction limits. When suspected outliers are confirmed (using methods discussed in the Unified Guidance), background data are flagged and deselected prior to constructing statistical limits.

If during detection monitoring, new site hydrogeologic information becomes available, the background data set may be reevaluated. The site will prepare an update to the background data set and calculate updated prediction limits for statistical comparison.

Natural systems continuously evolve due to physical and/or chemical changes in the environment. Periodic updating of background statistical limits is necessary to accommodate natural changes. For statistical comparisons, data for all wells and constituents may be re-evaluated when new data points are available (from the end of the previous background period) to determine whether earlier concentrations are representative of present-day groundwater quality. In some cases, earlier portions of the data may be deselected prior to construction of limits to make the limits more sensitive to changes in groundwater quality. Even though data may be excluded from the calculation, the values will continue to be reported and shown on data summaries, but the excluded data will be flagged to indicate that it is not being used.

2.3.1 Outlier Testing

An outlier is defined as an observation that is unlikely to have come from the same distribution as the rest of the data. The Unified Guidance directs that identified outliers should not be removed from the data set unless independent evidence of an error in the data exists. Under this direction, outliers are to be left in the data set and treated as “extreme values” during the statistical analysis. In some cases where the Sanitas™ database identifies outliers in the background groundwater quality data, professional judgment is required to determine whether the outlier significantly affects the statistical results to the extent that removal is deemed necessary, even when independent evidence of an error in the data does not exist.

In general, it is a common practice to remove obvious outliers from the database even when independent evidence does not exist. The removal of outliers tends to normalize the data and produces a more conservative statistical limit, since the data variability is decreased, thereby decreasing the standard deviation.

Outliers removed from the data set prior to the performance of prediction limit analyses are flagged in the Sanitas™ database with "(O)". The "(O)" flag in the Sanitas™ database denotes that the observation is an outlier and can be excluded from the statistical analysis. Suspected outliers at all wells for Appendix III and Appendix IV parameters are formally tested and, when identified, flagged in the database and deselected prior to construction of statistical limits.

Because nonparametric limits are set equal to the highest background observations, it is very important to carefully screen the background data for outliers and remove outliers that will result in false negative conclusions (i.e., an indication that no impact exists when a real impact is present).

2.3.2 Seasonality

Testing and adjusting data for seasonal factors ensures that seasonal effects will not confuse the test results. A seasonality test will be used to determine if significant seasonal effects are present when there are sufficient data to test for seasonality. When there are significant seasonal effects, then, where appropriate, the data will be de-seasonalized prior to constructing prediction limits. Background data will be tested again when there are at least four new values available to ensure that seasonal effects do not confound future analysis results. Seasonality adjustments will be used very cautiously, because true seasonality is not easily observed and generally requires a large amount of data gathered at very consistent sampling intervals.

2.3.3 Trend Analysis

Prior to the calculation of statistical limits, it is also important to perform a trend analysis on the background groundwater data. Even when the data are normally distributed, it is possible for trends to exist. If the background data are not inspected for trends, a pre-existing condition (i.e., related to either natural geochemical variability or potential contamination) may influence the statistical limit so that the limit is not representative of "true" background conditions. Because statistical analyses are most commonly performed relative to a prediction limit, positive trends in the data are usually more suspect than negative trends.

Trends observed in the background data from upgradient monitoring wells are typically treated differently than trends in downgradient monitoring well data. For instance, a trend in the concentration of a constituent in an upgradient monitoring well may indicate an off-site, upgradient source for the trend or some other source of natural variability (i.e., unrelated to CCR unit influence), while a trend in the concentration of a constituent in a downgradient monitoring well may indicate either variability due to an off-site, upgradient source or potentially a CCR unit influence. If upgradient and downgradient monitoring wells display similar trends in constituent concentrations, it is likely that the trends are due to natural variability or serial laboratory influences. If the downgradient monitoring wells display more extreme trends than the upgradient monitoring wells, it is more likely that there is evidence of a CCR unit influence which may need additional investigation.

Trends in the background groundwater monitoring data will be evaluated using a combination of the Mann-Kendall test (Hollander and Wolfe, 1973) and Sen's slope estimate (Gilbert, 1987) to estimate trends. Both the Mann-Kendall test and the Sen's slope estimate are non-parametric tests (i.e., not dependent upon data normality), which are not greatly influenced by large outliers in the data. The Mann-Kendall test assigns a score or test statistic to a series of data based on the negative or positive differences between consecutive data points. Sen's slope estimate calculates the "true" slope of a series of data. Sen's slope estimate is superior to simple linear regression, because it is not greatly influenced by outlying data points.

2.3.4 Sample Size Requirements

While prediction limit evaluations may be constructed with as little as four samples per well, the CCR Rule requires and the Unified Guidance recommends that a minimum of at least 8 independent background observations are collected for the purpose of constructing statistical limits. The reliability of the statistical results is greatly enhanced when the sample size is increased to eight or more. The sample size will increase organically during routine data collection. An increased sample size tends to more accurately characterize the variation and typically reduce the probability of erroneous conclusions. In addition, if a nonparametric prediction limit is required, the confidence level associated with the test will be dependent on the number of background data points available as well as the number of comparisons to the statistical limit.

2.3.5 Updating Background

Background data will be updated for interwell prediction limits by consolidating more recent sampling observations with historical background data during each sampling event. This will not only increase the background sample size but will also reduce the incidence of false positive results. Any new outliers in upgradient well data will be flagged and, if confirmed, not utilized in the construction of statistical limits.

When using intrawell methods, periodic updating of background data is recommended only after at least four new samples are available. Four new samples are available every two years; therefore, new sampling data may be consolidated with the background data every two years. Before updating the data for intrawell testing, it is necessary to verify that the most recent observations represent an unimpacted state. If the most recent data group is found to be statistically similar to the older data, the background data set may be updated. For updating of background data sets, the statistical test used to perform the comparison between two groups of data will be either an Analysis of Variance test or a Mann-Whitney/Wilcoxon Rank Sum test.

If during detection monitoring, new hydrogeologic information becomes available and it becomes evident that the background data set should be reevaluated, the site will prepare an update to the background data set and calculate updated prediction limits for statistical comparison.

2.3.6 Normality

After outlier removal, the data for each constituent are subjected to intrawell normality testing. The normality tests contained in the Sanitas™ database are used to determine whether the background data distributions are normal, transformed normal, or unknown. The Unified Guidance directs that the data be either normally or transform normally distributed before performing parametric prediction limit analysis.

For sample sizes of less than 50 samples, the Unified Guidance recommends the use of the Shapiro-Wilk Test of Normality. For sample sizes greater than 50 samples, the Unified Guidance recommends the use of the Shapiro-Francia Test of Normality. When background data cannot be normalized, nonparametric prediction limits will be calculated.

2.3.7 Handling Non-Detect Values

Simple substitution per the Unified Guidance will be used when non-detects comprise less than or equal to 15% of the individual well data. Simple substitution refers to the practice of substituting one-half the reporting or detection limit for non-detects. When the proportion of non-detects (NDs) in background falls between 16 and 50%, a non-detect adjustment such as the Kaplan-Meier or Regression on Order Statistics (ROS) method for adjustment of the mean and standard deviation will be used prior to constructing a parametric prediction limit. When the proportion of non-detects exceeds 50%, or when the data cannot be normalized, a nonparametric prediction limit will be used.

3.0 SITE-SPECIFIC STATISTICAL ANALYSIS METHOD

The following is a detailed description of the statistical analysis methodology that will be used for groundwater quality analysis at the AGS when monitored constituents are detected in any of the site monitoring wells.

Background sampling for the CCR program began in June 2016. A minimum of 8 background sampling events have been conducted for each Appendix III and Appendix IV constituent in each monitoring well.

For the statistical analysis of analytical results obtained from the existing monitoring well network, (1) the number of samples collected will be consistent with the appropriate statistical procedures recommended by the CCR Rule and the Unified Guidance; (2) the statistical method will comply with the USEPA-recommended performance standards; and (3) determination of whether or not there is an SSI over background values in the future will be completed per the above-mentioned regulations.

3.1 Detection Monitoring

Groundwater quality data will be evaluated through use of intrawell prediction limits, combined with a 1-of-2 resampling strategy for constituents listed in Appendix III of 40 CFR §257.

As previously demonstrated (Golder, 2018a and 2018b), the Willcox Playa is located hydraulically upgradient of the AGS. The regional water quality of the Willcox Playa consistently indicates elevated concentrations of total dissolved solids (TDS), chloride, sulfate, and bicarbonate. The highest concentration of those constituents is in the vicinity of the Willcox Playa where TDS concentrations are greater than 500 milligrams per liter (mg/L), chloride and sulfate concentrations are greater than 100 mg/L, and bicarbonate concentrations exceed 300 mg/L. Since the Willcox Playa is directly upgradient of the AGS, intrawell prediction limits are planned. Should increasing trends be observed in downgradient wells as a result of the non-CCR related upgradient source, it will be necessary to evaluate the statistical method for future comparisons.

The CCR rule does not currently allow for evaluating other constituents to replace Appendix III constituents that may be influenced by upgradient sources. Therefore, groundwater quality data will be evaluated through the use of intrawell prediction limits, combined with a 1-of-2 resampling strategy for Appendix III constituents. Intrawell prediction limits compare the most recent sample from a particular well to statistical limits constructed from historical background groundwater quality measurements at the same well. Since the AGS has a known upgradient source, the intrawell prediction limits may minimize the likelihood that spatial variability will contribute to invalid statistical limits. As described in Section 2.1.2, when groundwater quality in the upgradient monitoring wells displays spatial variability, intrawell statistical limits are more conservative (i.e., intrawell statistical limits will reveal apparent statistical exceedances prior to interwell statistical limits, because the standard deviation for an intrawell statistical limit is based on the variability of a single well instead of a pooled set of data from several background wells).

While it is noteworthy that statistical limits based on intrawell methods are likely to be more conservative (i.e., lower), it is also of note that the situation at AGS can result in false positives over time, since it is anticipated that upgradient groundwater will eventually make its way to downgradient locations, which may result in upward trends in the downgradient wells over time. In this case, the upward trends may not be indicative of influence from the CCR impoundment but may simply be a result of the migration and mixing of upgradient and downgradient groundwater over time. Verification sampling will help to reduce the incidence of false positives but will obviously not eliminate the effects of groundwater movement at the site.

If an SSI is identified, one independent resample may be collected to determine whether the initial SSI is verified. If the initial finding is not verified by resampling, the resampled value will replace the initial finding. If the resample confirms the initial finding, the SSI will be verified. The following statistical methods will be used.

3.1.1 Parametric Prediction Limit

These limits will be computed per the Unified Guidance when data can be normalized, possibly via transformation. The test alpha will be calculated based on the following configuration:

- Annual SWFPR = 0.10
- 1-of-2 resampling plan with a minimum of 8 background samples per well for prediction limit evaluation
- 1 of upgradient monitoring wells
- 3 of downgradient monitoring wells
- 7 constituents (Appendix III)

3.1.2 Nonparametric Prediction Limits

In general, the statistical power of a nonparametric test is significantly lower than the power of a parametric test, so parametric tests are preferred. If it is not possible to use parametric prediction limits, due to non-normality of the background data, a nonparametric limit will be used. The nonparametric prediction limit is simply the highest, historical background value. The associated confidence level takes into account the prospect of additional future compliance values (retests) when there is an initial, unverified SSI. The achieved confidence level is determined based on the background sample size, the number of monitoring wells in the network, and the number of proposed retests, using tables provided in the Unified Guidance.

3.1.3 Retesting Strategy

When the prediction limit analyses indicate an unverified SSI, one discrete verification resample from the indicating well(s) may be collected within 90 days and prior to the next regularly scheduled sampling event. In order for the test to be valid, the resample needs to be statistically independent which requires that sufficient time elapse between the initial sample and resample.

3.1.4 Responding to SSIs

If the statistical evaluation for an Appendix III constituent results in a verified SSI, the data must be evaluated to determine if the cause of the SSI is from a release from the CCR Unit or an alternative source. Possible alternative sources may include laboratory and/or sampling variability, errors in the statistical method or statistical evaluation, and/or natural variation. If the SSI can be attributed to one of these sources and the SSI was not caused by the CCR Unit, an Alternate Source Demonstration (ASD) can be completed. An ASD must be certified by a qualified professional engineer and completed in writing within 90 days of completing the statistical evaluation for a particular sampling event.

3.2 Assessment Monitoring

The following describes the general statistical procedures that will be used if a facility enters Assessment Monitoring as a result of SSIs during Detection Monitoring.

As described above, Assessment Monitoring may be initiated when there is an SSI over background in one or more wells for any of the Appendix III constituents. Wells are then sampled for constituents listed in Appendix IV of 40 CFR §257 within 90 days unless an ASD is made within that same timeframe.

When Appendix IV constituents are sampled, detected constituents are added to the list of constituents sampled semi-annually, and concentrations are compared to Groundwater Protection Standards (GWPS) using Confidence Intervals. The GWPS will be those specified in the applicable federal rules. Alternate contaminant levels (ACLs) may be established from upgradient wells when background groundwater concentrations are higher than the established Maximum Contaminant Levels (MCLs) or federally derived risk-based screening levels (RBSLs). On an annual basis, all Appendix IV constituents will be sampled, and detected constituents added to the list of constituents sampled semi-annually.

Parametric confidence intervals around the population mean will be constructed at the 99% confidence level when data follow a normal distribution, and around the geometric mean (or population median) when data follow a transformed-normal distribution.

Non-parametric confidence intervals will be constructed when data do not pass a normality test and cannot be normalized via transformation. The confidence level associated with the non-parametric tests is dependent on the number of values used to construct the interval. Confidence intervals require a minimum of four samples; however, eight samples are recommended. When a well/constituent pair does not have the minimum sample requirement, the well/constituent pair will continue to be reported and tracked using time series plots and/or trend tests until such time that enough data are available.

In Assessment Monitoring, a well is determined to be out of statistical compliance when the Lower Confidence Limit (LCL), or the entire confidence interval, is higher than the GWPS as discussed in the Unified Guidance. Unless an alternate source is demonstrated, assessment of corrective measures is initiated at that time and remediation efforts will continue to be evaluated through the use of Confidence Intervals to determine the effectiveness of remediation. As described in the CCR rule, if constituents are detected at a statistical level higher than a GWPS during Assessment Monitoring, it is referred to as a Statistically Significant Level (SSL).

3.3 Corrective Action

Similar to assessment monitoring, the following describes the general statistical procedures that will be used if a facility enters Corrective Action monitoring due to SSL during Assessment Monitoring.

Once corrective measures have been initiated, semi-annual sampling will continue, and Confidence Intervals will monitor the progress of remediation efforts. As in Assessment Monitoring, Confidence Intervals are compared to established GWPS to determine when clean-up levels are achieved. However, in Corrective Action, a well/constituent pair is declared to no longer be an SSL when the entire interval falls below a specified limit (i.e., the Upper Confidence Limit [UCL] falls below the limit). Alternatively, compliance is achieved when the LCL of the Appendix IV constituents is no longer higher than the GWPS for a period of three consecutive years.

4.0 STATISTICAL METHOD CERTIFICATION

USEPA's "Disposal of Coal Combustion Residuals from Electric Utilities" Final Rule (40 CFR Part 257 and Part 261), §257.93, requires the owner or operator of an existing CCR unit to identify a statistical method to be used in evaluating groundwater monitoring data for each specified constituent. The owner or operator must obtain a certification from a qualified professional engineer stating that the selected statistical method is appropriate for evaluating the groundwater monitoring data for the CCR management area meeting the requirements of 40 CFR §257.93. 40 CFR §257.93(f)(6) states the following:

> *"The owner or operator of the CCR unit must obtain a certification from a qualified professional engineer stating that the selected statistical method is appropriate for evaluating the groundwater monitoring data for the CCR management area. The certification must include a narrative description of the statistical method selected to evaluate the groundwater monitoring data."*

Statistical Methodology

The selected statistical method for the CCR unit at AGS was developed in accordance with 40 CFR §257.93(f) using methodology presented in the Unified Guidance.

The statistical test used to evaluate the groundwater monitoring data from the certified monitoring well network will be either the prediction limit or tolerance interval method. Either interwell or intrawell statistical methods will be used. The current data from the upgradient wells will be evaluated to determine if the Willcox Playa (upgradient of the AGS) continues to affect the statistical evaluation.

Interwell statistical methods are conducted by evaluating the data from downgradient wells and comparing the data to upgradient background groundwater quality. Using this approach, background data from upgradient well(s) will be pooled to calculate a PL for appropriate constituents. Data from the downgradient monitoring wells will be evaluated by comparing individual results to the PL following each monitoring event. An "unverified SSI" occurs when any downgradient well data are higher than the upgradient PL.

Intrawell statistical methods compare the most recent sample from a particular well to statistical limits constructed from historical background groundwater quality measurements at the same well. Since the AGS has a known upgradient source, the intrawell PL may minimize the likelihood that special variability will contribute to invalid statistical limits.

If data from a sampling event initially are higher than the PL, a resampling strategy for applicable constituents may be used to verify the result. In resampling, independent resamples will be collected and evaluated to determine whether the initial unverified SSI is verified. If the resample result does not verify the initial result, the initial unverified SSI is considered a spurious result and the resample value will replace the initial result. When the resample confirms the initial finding, an SSI is verified. An SSI is determined only if the resample verifies the initial unverified SSI (i.e. the resample also exceeds the PL) or if AEPCO elects not to resample to confirm the initial finding.

Statistical Certification

I hereby certify that, in accordance with the requirements of 40 CFR §257.93, the selected statistical method as described above is appropriate for evaluating the groundwater monitoring data for the CCR Units located at AEPCO's AGS located at 3525 N. Highway 191, Cochise, Arizona 85606, and designated as Ash Pond Unit 1 – 4 Multi-Unit and Scrubber Pond No. 2 Single Unit.

Golder Associates Inc.

Dawn L. Prell, CPG
Senior Consultant

Kent R. Johnejack, PE
Arizona Registered Professional Engineer

dlp/scp/krj

p:\2019 projects\19116278 ags stat analysis plan\kj2_aepco-ags stats plan update 5.21.2019_final_6.6.2019 (002).docx

5.0 REFERENCES

Gibbons, Robert Ltd. And Discerning Systems, Inc., 2000. Statistical Guide

Gilbert, R.O., 1987. Statistical Methods for Environmental Monitoring

Golder, 2017. Statistical Analysis Plan, Apache Generating Station, Cochise County, Arizona, Prepared for the Arizona Electric Power Cooperative, Inc., October 17, 2017

Golder, 2018a. Alternate Source Demonstration, Apache Generating Station Ash Pond 1-4, Multi-Unit System. Prepared for Arizona Electric Power Cooperative, Inc., December 28, 2018

Golder, 2018b. Alternate Source Demonstration, Apache Generating Station Scrubber Pond No. 2, Single-Unit System. Prepared for Arizona Electric Power Cooperative, Inc., December 28, 2018

Hollander, M. and Wolfe, D.A., 1973. Nonparametric Statistical Methods

USEPA, 2009. Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities, Unified Guidance, Office of Solid Waste EPA/530/R/09/007, March 2009

golder.com