REPORT

STATISTICAL ANALYSIS PLAN

Grand Haven Board of Light and Power

Submitted To: Grand Haven Board of Light and Power
17000 Eaton Drive
Grand Haven, MI 49417

Submitted By: Golder Associates Inc.
27200 Haggerty Road, Suite B-12
Farmington Hills, MI 48331-5719

Distribution: Mr. Paul Cederquist, GHBLP
Grand Haven Board of Light and Power-Operating Record
Golder Associates Inc.

October 17, 2017 **1775416B**

Golder Associates

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 during sampling events and identify statistically significant increases (SSIs) in concentrations above background levels, if applicable. 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 developed based on site specific review of the data to develop an appropriate analysis of the groundwater data collected for compliance with the CCR rules. This SAP provides a description of the statistical approach and methods used in accordance with the CCR rule reporting requirements [40 CFR 257.93(f)(6)]. .

This SAP describes statistical procedures for the CCR units (Ash Pond A&B). Specifically the SAP describes:

- The criteria a facility's background data must meet to justify the use of the selected methodologies
- A description of the statistical methodologies 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.

2.0 STATISTICAL APPROACH FOR DETECTION MONITORING PROGRAMS

The statistical methodology will meet the criteria referenced in the CCR Rule as well as the USEPA Unified Guidance.

2.1 Prediction Limits

Prediction limit tests will be used for Appendix III constituents to represent *current* conditions. The prediction limit method is preferred over control charts and tolerance limits. Relative to control charts, prediction limits generally result in a more conservative statistical limit and are generally more straightforward. Prediction limits are also preferred over tolerance limits, since tolerance limits are calculated to include a specified percentage of background observations (e.g., typically 99 percent), 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 compares groundwater quality from upgradient well data to construct statistical limits which are used to evaluate whether downgradient wells are in compliance. 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 statistically similar to downgradient groundwater (assuming no release)

Initially, newly regulated facilities have 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 semiannual groundwater monitoring.

The USEPA Unified Guidance recommends switching from interwell methods to intrawell methods when it can be reasonably demonstrated that no contamination from current practices or waste management at the regulated facility is present for well/constituent pairs in Detection Monitoring.

2.1.2 Intrawell Prediction Limits

Intrawell Prediction Limits, which compare the most recent sample from a particular well to statistical limits constructed 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 statistical limits (i.e., more conservative) and result in fewer false positive occurrences as long as there has been no previous release. According to the USEPA Unified Guidance, “One key advantage (for intrawell testing) is confounding results due to spatial variability are eliminated, since all data used in an intrawell test are obtained from a single location.

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). In particular, 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.

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

The USEPA 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 semiannual 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 Reference Power Curves following this report demonstrate the statistical power for the methods selected at this site. The USEPA recommend statistical power of approximately 55% at 3 standard deviations or 80% at 4 standard deviations over the background mean.

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 though no impact has occurred. 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 five percent 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. It was already demonstrated above that the number of comparisons should be limited by limiting the number of constituents that are statistically analyzed on a semiannual basis.

Decreasing the per test FPR will decrease the SWFPR. However, the U.S. Environmental Protection Agency mandates that the per test FPR must be at least one percent (i.e., minimum of a 99 percent 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.

2.2.2 Statistical Power

In addition to the SWFPR, the USEPA 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 upgradient well data for the purpose of constructing statistical limits, it is necessary to screen prospective background data. Data are screened for anomalies such as high outliers that will artificially elevate prediction limits, and for trends over time. Time series plots, box plots and statistical outlier tests can be reviewed to identify anomalous observations. Suspected observations are de-selected from prospective background data prior to establishing prediction limits. When suspected outliers are confirmed through methods as discussed in the USEPA Unified Guidance, background data are flagged and deselected prior to constructing statistical limits. Any values flagged as outliers will be summarized in the report.

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.

2.3.1 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 Sanitas™ identified outliers in the background groundwater quality data, professional judgment was used to determine whether the outlier significantly affected the statistical results to the extent that removal was deemed necessary, even when independent evidence of an error in the data did not exist.

In general, it is a good 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 therefore provides a more appropriate statistical limit. Outlier removal also tends to produce a more conservative statistical limit, since the data variability is decreased, thereby decreasing the standard deviation.

Outliers may be removed from the data set prior to the performance of prediction limit analyses are flagged with “(O)”. The “(O)” flag in the Sanitas™ database denotes that the observation is an outlier and can be excluded from the statistical analysis.

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.

2.3.3 Trend Analysis

Prior to the calculation of statistical limits, it is also important to perform trend analysis on the background groundwater monitoring 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 potentially 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. A positive trend in an Appendix III constituent has the potential to be related to a CCR unit and generally needs additional investigation.

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. 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 use the Sanitas™ statistical software. Sanitas™ uses 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 and the USEPA Unified Guidance recommends that a minimum of at least 8 independent background observations be collected for the purpose of constructing statistical limits. The reliability of the statistical results is greatly enhanced by increasing the sample size to eight or more. An increased sample size tends to more accurately characterize the variation and typically reduce the probability of erroneous conclusions. Furthermore, if a nonparametric prediction limit is required, the confidence level associated with the test will be dependent on the number of background data 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 not found to be statistically different than the older data, the background data may be updated.

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 were subjected to intrawell normality testing. The normality tests contained in Sanitas™ 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 USEPA 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 Grand Haven BLP when monitored constituents are present in any of the downgradient wells.

Background sampling for the CCR program began in March 2017. A total of 8 background sampling events have been conducted.

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 as recommended by the CCR Rule and the USEPA Unified Guidance; (2) the statistical method will comply with the USEPA-recommended performance standards; and (3) determination of whether or not there is a statistically significant increase (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 interwell prediction limits, combined with a 1-of-2 resampling strategy for Appendix III constituents (boron, calcium, chloride, fluoride, pH, sulfate, and total dissolved solids (TDS)). If a statistical exceedance is found, one independent resample will be collected to determine whether the initial exceedance is verified.

If the initial finding is not verified by resampling, the resampled value will replace the initial finding. When the resample confirms the initial finding, the exceedance will be reported. The following statistical methods will be used:

3.1.1 Parametric Prediction Limit

These limits will be computed per USEPA 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 interwell prediction limits
- # of upgradient monitoring wells
- # of downgradient monitoring wells

- 7 constituents (Appendix III)

3.1.2 Nonparametric Prediction Limits

The highest background value will be used to set the upper nonparametric prediction limit. The associated confidence level takes into account the prospect of additional future compliance values (retests) when there is an initial exceedance. 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 USEPA Unified Guidance.

3.1.3 Retesting Strategy

When the prediction limit analyses indicates an initial exceedance, one discrete verification resample from the indicating well(s) will 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. A minimum time interval between samples will be established to ensure that separate volumes of groundwater are being sampled.

3.2 Assessment Monitoring

The following describes the general statistical procedures that will be used if a facility enters Assessment monitoring because of SSIs in the detection monitoring program. Site-specific and event-specific SAPs may be developed at that time according to permit or regulatory requirements.

Assessment Monitoring may be initiated when there is a confirmed SSI over background in one or more wells for any of the Appendix III constituents. Wells are then sampled for Appendix IV constituents within 90 days unless a demonstration is made within that same timeframe that the statistical exceedance resulted from a source other than the CCR unit, or an error in sampling, analysis, statistical evaluation, or natural variation in groundwater.

When Appendix IV constituents are sampled, detected constituents are added to the list of constituents sampled semiannually, and concentrations are compared to Groundwater Protection Standards (GWPS) through the use of Confidence Intervals. The GWPS will be those specified in applicable state or federal rules. Alternate contaminant levels (ACLs) may be established from upgradient wells when groundwater concentrations upgradient of the facility are higher than the established MCLs. On an annual basis, all Appendix IV constituents will be sampled and newly detected constituents added to the list of constituents sampled semiannually.

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 a 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 compliance when the Lower Confidence Limit (LCL), or the entire interval, exceeds the GWPS as discussed in the USEPA Unified Guidance. Assessment of corrective action 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.

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 because of Statistically Significant Increases (SSIs) in the Detection monitoring program. Site-specific and event-specific SAPs may be developed at that time according to permit or regulatory requirements.

Once remediation activities begin, semiannual sampling will continue and Confidence Intervals will monitor the progress of remediation efforts. Confidence Intervals are compared to GWPS, to determine when clean-up levels are achieved.

In Corrective Action a well/constituent pair is declared to no longer be an SSI over the GWPS 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 Lower Confidence Limit (LCL) of the Appendix IV constituents does not exceed 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

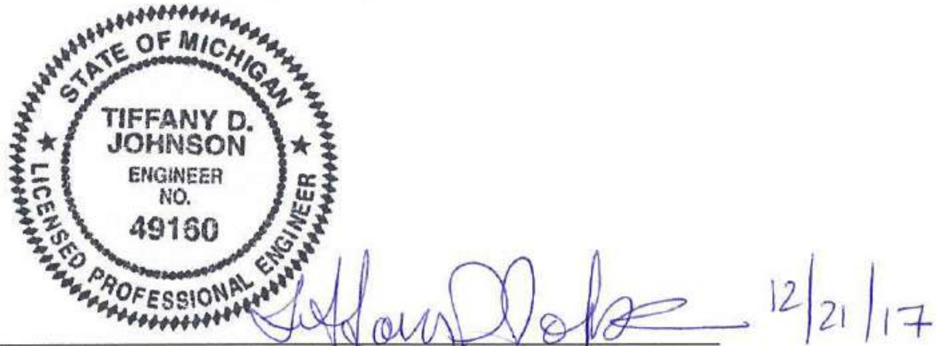
The selected statistical method for Grand Haven Board of Light and Power Ash Pond (AP-A&B) was developed in accordance with 40 CFR §257.93(f) using methodology presented in *Statistical Analysis of Groundwater Data at RCRA Facilities, Unified Guidance*, March 2009, USEPA 530/R-09-007 (Unified Guidance).

The statistical test used to evaluate the groundwater monitoring data will be the prediction interval method. Interwell statistical methods will be used as specified in the Statistical Analysis Plan – meaning that data from downgradient wells will be compared to upgradient background groundwater quality. Using this approach, background data from the network of upgradient wells will be pooled to calculate a Prediction Limit (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 "initial exceedance" occurs when any downgradient well data exceed the upgradient PL.

If data from a sampling event initially exceed the interwell PL, a resampling strategy for applicable constituents will be used to verify the result. In resampling, independent resamples will be collected and evaluated to determine whether the initial exceedance is verified. If the resample result does not verify the initial result, the initial exceedance is considered a spurious result and the resample value will replace the initial result. When the resample confirms the initial finding, a statistically significant increase (SSI) is declared. An SSI is determined only if the resample verifies the initial exceedance (i.e. the resample also exceeds the PL).

CERTIFICATION

I hereby certify that the groundwater statistical plan for the Grand Haven Board of Light and Power Ash Ponds (AP-A&B) located at 1231 North Third Street in Grand Haven, Michigan, has been developed in accordance with 40 CFR §257.93(f).



Tiffany D. Johnson, P.E.
Licensed State of MI, PE No. 6201049160
12/21/2017

5.0 REFERENCES

Criteria for Classification of Solid Waste Disposal Facilities and Practices. 40 CFR §257. (2016)

Electric Power Research Institute. 2015. Groundwater Monitoring Guidance for the Coal Combustion Residuals Rule. Palo Alto, CA. 3003006287

Environmental Resources Management Michigan, Inc. 2017. Sampling and Analysis Plan for the Grand Haven Board of Light and Power.

United States Environmental Protection Agency. 2009. Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities: Unified Guidance. USEPA 530/R-09-007.

Established in 1960, Golder Associates is a global, employee-owned organization that helps clients find sustainable solutions to the challenges of finite resources, energy and water supply and management, waste management, urbanization, and climate change. We provide a wide range of independent consulting, design, and construction services in our specialist areas of earth, environment, and energy. By building strong relationships and meeting the needs of clients, our people have created one of the most trusted professional services organizations in the world.

Africa	+ 27 11 254 4800
Asia	+ 852 2562 3658
Australasia	+ 61 3 8862 3500
Europe	+ 356 21 42 30 20
North America	+ 1 800 275 3281
South America	+ 56 2 2616 2000

solutions@golder.com
www.golder.com

Golder Associates Inc.
27200 Haggerty Road, Suite B-12
Farmington Hills, MI 48331-5719
Tel: (248) 295-0135
Fax: (248) 295-0133

Golder Associates

Engineering Earth's Development, Preserving Earth's Integrity