CCR COMPLIANCE
STATISTICAL METHOD FOR GROUNDWATER DATA EVALUATION
ASH FILTER PONDS AND ASH DISPOSAL SITE

Prepared for:

GenOn Northeast Management Company
Keystone Generating Station
Shelocta, Pennsylvania

Prepared by:

Aptim Environmental & Infrastructure, Inc.
Pittsburgh, Pennsylvania

October 2017

Table of Contents

List of Figures

1.0 Introduction

Title 40 Code of Federal Regulations §257.91 requires the owner or operator of Coal Combustion Residuals (CCR) landfills and surface impoundments, also known as CCR units, to implement a groundwater monitoring system. These requirements are part of the overall CCR Rule (or Rule) which was published in the Federal Register on April 17, 2015 and which became effective on October 19, 2015. The referenced groundwater monitoring system for each defined CCR unit must consist of a sufficient number of wells (minimum one upgradient and three downgradient) installed at appropriate locations to accurately determine background groundwater quality and also to accurately represent the quality of groundwater passing the boundary of the CCR unit.

Further, §257.93(a) of the Rule requires that a groundwater sampling and analysis program be established to include consistent procedures to ensure that the monitoring results accurately represent the quality of groundwater at the upgradient and downgradient wells. In addition, §257.93(f) also requires selection of a statistical method for use in determining if a statistically significant increase over background concentrations in groundwater has occurred at one or more of the downgradient monitoring well locations. Candidate statistical methods are outlined in §257.93(f)(1-5) and corresponding performance standards (dependent upon the method selected) are specified in §257.93(g)(1-6). Lastly, §257.93(f)(6) requires the owner or operator of the CCR unit to obtain a certification from a 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.

The Keystone Generating Station, operated by GenOn Northeast Management Company, is a coal-fired power plant located in Shelocta, Pennsylvania. The Rule applies to this facility due to the management/disposal of CCR materials that are generated from the combustion of coal. CCR units associated with station operations include the Keystone Ash Disposal Site (represented by the East Valley and West Valley Disposal Sites), and three Ash Filter Ponds (Ponds “A”, “B”, and “C”) used for the management of bottom ash. Each of these CCR units has a dedicated groundwater monitoring well network that meets the requirements of §257.91 with regard to number and appropriate locations of wells (certification provided under separate cover). Additionally and in accordance with the provisions of §257.91(d) of the Rule, the monitoring network for the Ash Filter Ponds has been established to provide coverage in the context of a multiunit system, encompassing all three ponds (Ponds “A,” “B,” and “C”) collectively.

This Certification has been prepared to comply with the requirements of §257.93(f)(6), addressing the statistical method selection for the CCR units noted above. This Certification will be placed in the Keystone facility's operating record per §257.105(h)(4), noticed to the State Director per §257.106(h)(3), and posted to the publicly accessible internet site per §257.107(h)(3).

2.0 Statistical Method Selection and Background Data Evaluation

As previously noted, each of the station's CCR units has a dedicated groundwater monitoring system, represented by at least one upgradient/background well and a minimum of three downgradient wells. Specific to the Ash Filter Ponds and as shown on Figure 1, the groundwater monitoring system includes upgradient well MW-5, and downgradient wells MW-6, MP-29, and MP-30. Specific to the Ash Disposal Site and as shown on Figure 2, the groundwater monitoring system for the East Valley area is represented by background well MP-21 and downgradient wells MP-4, MP-17B and MP-18. Similarly, the groundwater monitoring system for the West Valley area is represented by background well MP-21 and downgradient wells MP16, MP-23, and MP-24. As indicated, the Keystone Ash Disposal Site consists of the contiguous East Valley and West Valley areas and is located north of the Station proper. East Valley was constructed in two stages, including Stage I (closed) and Stage II (currently active). West Valley was constructed to initially include Stage III (currently active), and is currently in the midst of a permitted expansion (Stage IV) that has been underway since before the effective date of the CCR Rule.

Choosing an appropriate statistical method is paramount in developing a sound and defensible groundwater monitoring program. As such, the statistical method should be commensurate with knowledge of the basic site-specific characteristics such as number and configuration of wells, the water quality constituents being measured, and general hydrology. The method should also be selected with reference to the statistical characteristics of the monitored parameters such as proportion of non-detects, type of concentration distribution (e.g., normal, lognormal) and presence or absence of spatial variability.

For each of the CCR units at the Keystone Station, an interwell prediction limit approach has been selected. In addition to being one of the candidate methods cited under §257.93(f)(1-5), the interwell prediction limit method is among those recommended in U.S. EPA's (EPA) Unified Guidance document ("Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities," March 2009). This guidance document was developed in order to assist the EPA and the regulated community in testing and evaluating groundwater monitoring data under 40 CFR §258, §264, and §265 (relating to solid waste and hazardous waste management facilities).

The prediction limit approach is flexible and conforms to varying data distributions, frequencies of non-detects, and whether or not the data exhibit a significant trend. Parametric tests are used for those datasets which follow a known and identifiable distribution, with the most common examples in groundwater monitoring being the normal and the lognormal. If a specific distribution cannot be determined, non-parametric test methods can be used. Non-parametric

tests do not require a known statistical distribution and can be helpful when the data contain a substantial proportion of non-detects.

Prediction limits are generally easy to construct and have a straightforward interpretation. Only background values are used to construct a concentration-based prediction limit, which is then compared to one or more future observations from a compliance point population (e.g., downgradient compliance wells). For purposes of detection monitoring (§257.94 of the Rule) and potential assessment monitoring (§257.95 of the Rule) and as is typical, a one-sided upper prediction limit will be constructed from the background data. A noted exception is pH which will also have a lower-prediction limit, essentially creating a range of values deemed representative of background. Specific to the Keystone CCR units and further detailed below, separate prediction limits will be constructed for the Ash Filter Ponds and the Ash Disposal Site, covering each of the relevant Appendix III and IV constituents.

2.1 Establishment of Background Groundwater Values

Based on the groundwater data collected from each upgradient well (corresponding to the minimum eight required rounds of sampling per §257.94[b]), prediction limits of background for each CCR unit will be established for the constituents listed in Appendices III and IV, as follows:

Appendix III	*Appendix IV*
Total Boron	*Total Antimony*
Total Calcium	*Total Arsenic*
Chloride	*Total Barium*
Fluoride	*Total Beryllium*
Total Dissolved Solids	*Total Cadmium*
Sulfate	*Total Chromium*
pH	*Total Cobalt*
---	*Fluoride*
---	*Total Lead*
---	*Total Lithium*
---	*Total Mercury*
---	*Total Molybdenum*
---	*Total Selenium*
---	*Total Thallium*
---	*Radium 226 + 228*

2.1.1 Outlier Testing

Prior to use in establishment of the prediction limits, the background datasets (on a constituent by constituent basis) will be evaluated for potential outliers. However, in this regard, EPA's Unified Guidance recommends that statistical outliers should not be removed or altered unless independent evidence of an error exists. Accordingly, if evidence of an error is found to exist, these points will be removed from the dataset prior to calculation of the prediction limits.

2.1.2 Spatial Variability Analysis

Spatial variability in groundwater monitoring is generally understood to be present when the mean levels of a given constituent vary from one well to the next. For situations in which more than one upgradient well exists, (such is not the case for any of the Keystone CCR units) the data from these wells will be reviewed for evidence of statistically significant spatial variability based on an Analysis of Variance (ANOVA) test. If significant spatial variation is identified, consideration may be given to modification of the approach, including potential transition to an intrawell method for future comparisons at the compliance wells (assuming that it can be established convincingly that they have not been impacted by the CCR unit).

2.1.3 Temporal Variability Analysis

Temporal variability in groundwater monitoring exists when the distribution of measurements varies with the times at which sampling or analytical measurement occurs. There are several reasons that temporal variability can occur, with the most common being seasonal fluctuations. In the event that seasonality is detected, the data can be "de-seasonalized." However, corrections for seasonality are to be applied cautiously, as they represent extrapolation into the future. There should be a defensible physical explanation along with sufficient empirical evidence for seasonality before corrections are made. Any adjustments made for temporal variability would be done as described within the Unified Guidance. With respect to temporal variability, it is emphasized that clear identification of any potential trends/fluctuation would be limited until several additional years of data are collected beyond the first eight rounds of background sampling.

2.1.4 Determination of Data Distribution

Determining the distribution of data (normal vs. non-normal) is important since it forms the basic premise for parametric tests. For a normal distribution, this means that the density of the data or the natural log of the data follows the traditional bell-shaped curve, with the greatest number of values being centered around the mean and fewer values being a significant distance from the mean.

Normality will first be evaluated using the Shapiro-Wilk Test with a specified Alpha (α) of 95 percent (interwell prediction limit default). The Shapiro-Wilk Test is based on the assumption that if a dataset (or the natural logs of the dataset) is normally distributed, then the ordered values should be highly correlated with the corresponding quantiles of the normal distribution. The Shapiro-Wilk test statistic, W, will be large when the probability plot of the data indicates a straight line, but will be small if there are significant bends or curves in the plotted data. The test statistic will be compared to published critical values, and the assumption of normality rejected when the calculated test statistic falls below the critical values.

The denominator, d, of the W test statistic calculation is computed as follows (Gilbert, 1987):

$$d = \sum_{i=1}^{n} (X_i - \bar{X})^2 = \sum_{i=1}^{n} {X_i}^2 - \frac{1}{n}\left[\sum_{i=1}^{n} X_i\right]^2$$

Where:

X_i = the i^{th} smallest ordered value in the sample,
$\bar{X}$ = the mean of the n observations, and
n = the number of observations.

The observations are then ordered from smallest to largest and k is computed where:

$$k = \frac{n}{2} \quad \text{(if } n \text{ is even), and}$$
$$k = \frac{n-1}{2} \quad \text{(if } n \text{ is odd)}$$

The W test statistic is then computed as follows:

$$W = \frac{1}{d}\left[\sum_{i=1}^{k} a_i\left(X_{[n-i+1]} - X_i\right)\right]^2$$

Where:

$a_1, a_2, \ldots, a_k$ are provided in Table A6 (Gilbert, 1987).

The data will be tested at the α=0.05 significance level, with the null hypothesis being rejected if W is less than the quantile given in Table A7 (Gilbert, 1987).

However, if the original dataset fails the test, there are a series of transformations that may be applied to determine if any helps fit the data to the bell-shaped curve. The Ladder of Powers (Helsel and Hirsch, 1992) includes the following transformations in order of execution: x, $x^{1/2}$, x^2, $x^{1/3}$, x^3, ln(x), x^4, x^5, x^6. If one or more of these transformations passes the normality test, all

data will be transformed prior to the construction of any prediction limits; the data transformation that best normalizes the distribution will be used.

Specialized software (obtained from Sanitas™) will be utilized to aid in performing the above transformations, in addition to other statistical evaluations, including ultimate calculation of the background prediction limits. This software relies on a decision-logic framework that progresses through a series of statistical step-flow charts and testing algorithms, arriving at the best suited application and making any necessary adjustments or transformations to the datasets.

2.1.5 Managing Non-Detects

As is commonplace in groundwater monitoring programs and in part due to natural variability, measurable levels of constituents may be detected during certain sampling events and then be absent (non-detect) during other events. In preparation for determining the distribution of the background datasets (described above in Section 2.1.4), the following recommended procedures will be adopted to manage non-detects:

100 Percent Non-Detects. If 100 percent of the analyses for a constituent resulted in non-detects at or below the reporting limit, it will be assumed that the constituent is not present and no further statistical evaluation will be performed. The practical quantitation limit or method detection limit will then be assumed to be the upper prediction limit.

90 to 100 Percent Non-Detects. If 90 to 100 percent of the analyses for a constituent results in non-detects at or below the reporting limit, a non-parametric evaluation will be used wherein the highest detected concentration will serve as the upper prediction limit.

50 to 90 Percent Non-Detects. If 50 to 90 percent of the analyses for a constituent result in non-detects at or below the reporting limit, these values will be replaced with one half the reporting limit and a nonparametric confidence interval will be constructed wherein the highest detected concentration is utilized as the upper prediction limit.

15 to 50 Percent Non-Detects. If 15 to 50 percent of the analyses for a constituent result in non-detects at or below the reporting limit, the detected values will be evaluated using either a parametric or non-parametric method commensurate with published guidance.

0 to 15 Percent Non-Detects. If 0 to 15 percent of the analyses for a constituent results in non-detects at or below the reporting limit, these values will be replaced with one half the reporting limit and the data tested for normality.

If the data are normally or lognormally distributed, the sample mean and sample standard deviation will be adjusted using Aitchison's method, and a parametric evaluation (Section 2.1.6.1) will be performed to determine the upper prediction limit. If the data are not normally or

lognormally distributed, a non-parametric method (Section 2.1.6.2) will be utilized wherein the highest detected concentration for each constituent will serve as the upper prediction limit.

2.1.6 Parametric and Non-Parametric Evaluations

2.1.6.1 Parametric Evaluation

The parametric evaluation of normally and lognormally distributed data with 50 percent or fewer non-detects will be performed according to the methods described in the Unified Guidance. The 95 percent prediction limit will be calculated assuming that one sample would be taken from one well during two future sampling periods (one sampling period and one resampling event if necessary to confirm any observed exceedance). The equation for the 95 percent prediction limit is given by:

$$95\%\ Prediction\ Limit\ = \bar{x} + t_{1-0.05/m,n-1} s \sqrt{1 + \frac{1}{n}}$$

Where:

$\bar{x}$ = the sample mean of the detected or adjusted results
S = sample standard deviation of the detected or adjusted results
$t_{1-0.05/m,n-1}$ = the students t-coefficient for degrees of freedom (n-1) and confidence level (1-0.05/m)
n = the number of samples
m = the number of future samples

For this analysis, $\bar{x}_s$ and S_s are calculated as:

$$\bar{x}_s = \frac{1}{n}\sum_{i=1}^{n} x_i$$

and,

$$S_s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\bar{x}_s - x_i)^{1/2}}\ .$$

As described above, prediction limits can be constructed to accurately account for the number of tests to be conducted (a resampling plan), so as to limit the site-wide false positive rate and ensure an adequate level of statistical power. The Unified Guidance suggests that the annual site-wide false positive rate be no greater than 10 percent (i.e. 5 percent per semiannual event; 2.5 percent per quarterly event).

The basic equation for estimating the site-wide false positive rate (not including resampling) is the following:

$$\alpha_{cum} = 1 - (1 - \alpha_{test})^{n_T}$$

Where:

α_{cum} = site-wide false positive rate
α_{test} = test-wide false positive rate
n_T = number of wells x number of constituents in a calendar year

By rearranging to solve for α_{test}, the 10 percent design site-wide false positive rate (0.1) can be substituted for α_{cum} and the needed per-test false positive error rate calculated as:

$$\alpha_{test} = 1 - (1 - 0.9)^{1/n_T}$$

Aitchison's Adjustment

Aitchison's method adjusts the sample mean and sample standard deviation to account for non-detects below the reporting limit in data that are normally or lognormally distributed and have between 15 and 50 percent non-detects. Aitchison's method assumes that non-detect samples do not contain the constituent of concern, are free of contamination, and could be considered as having a zero concentration in the analysis.

Using the data above the detection level, the sample mean and sample variance are calculated as follows:

$$\bar{X}_d = \sum_{i=1}^{m} X_i$$

and,

$$s_d^2 = \frac{1}{m-1}\left\{\sum_{i=1}^{m} X_i^2 - \frac{1}{m}\left(\sum_{i=1}^{m} X_i\right)^2\right\}$$

The adjusted sample mean and sample variance are then calculated as follows:

$$\bar{X} = \left(\frac{m}{n}\right)\bar{x}_d$$

and,

$$s^2 = \frac{(m-1)}{(n-1)}s_d^2 + \frac{m(m-1)}{n(n-1)}\bar{X}_d^2$$

Where:

m = the number of detects,
n = the total number of samples.

2.1.6.2 Non-Parametric Evaluation

A non-parametric evaluation is one that is not based upon specific parameters of the variate, such as the sample mean and sample standard deviation. A non-parametric evaluation will be used when data do not follow a distribution that can be predicted according to statistical parameters, or in those instances where a large proportion of the samples are reported as non-detects (i.e., greater than 90 percent). The non-parametric evaluation will take the highest detected concentration as the upper prediction limit for the constituent.

3.0 Detection and Assessment Monitoring

3.1 Detection Monitoring

Per §257.90(b)(1)(iii-iv) of the Rule and no later than October 17, 2017, groundwater detection monitoring for existing CCR units is to have included performance of eight rounds (at a minimum) of background sampling, and the corresponding start of evaluation for statistically significant increases over background with regard to the Appendix III constituents. Accordingly, the data generated from the eight rounds of background sampling will be subjected to the statistical protocols outlined in Section 2.0, and upper prediction limits established for each Appendix III constituent (pH will also have a lower prediction limit). To support the evaluation of statistically significant increases, samples from the groundwater wells associated with each of the Keystone CCR units will be collected on a semiannual frequency (per §257.94[b]) and analyzed for the Appendix III constituents. The data from the downgradient wells at each unit will then be compared to the upper prediction limits on a constituent by constituent basis.

If during the course of semiannual detection monitoring an Appendix III constituent (in any of the downgradient wells) is measured above its respective upper prediction limit (or below the lower prediction limit in the case of pH), this finding will constitute a preliminarily identified statistically significant increase. Pursuant to this finding and within 90 days, a repeat sampling event will be conducted and further efforts undertaken to determine if possible laboratory error or some other confounding condition has been noted, or if an alternate source (other than the CCR unit) could be responsible for the increase. If these efforts do not provide the ability/evidence to either nullify the increase or delineate an alternate source, then the affected CCR units will transition from detection monitoring to assessment monitoring.

3.2 Assessment Monitoring

As described above, if a statistically significant increase is confirmed (and cannot be dismissed or alternate source identified) then the affected CCR unit must move from detection monitoring to assessment monitoring (§257.95 of the Rule). Notice of this transition must be placed in the facility operating record per §257.105(h)(5)94(e)(3), and appropriate notification made to the State Director per §257.106(h)(4) along with posting to the publicly accessible internet site per §257.107(h)(4).

Within 90 days of entry into assessment monitoring, all wells associated with the affected CCR unit will be sampled for the list of Appendix IV constituents. Subsequently, and within 90 days of obtaining the results from the initial round of sampling and on at least a semiannual basis thereafter, all wells will be analyzed for the constituents in Appendix III and for those constituents in Appendix IV that were detected.

In similar fashion to that described above in Section 3.1, the data from the eight rounds of background sampling will be subjected to the statistical protocols outlined in Section 2.0 and upper prediction limits established for each Appendix IV constituent. In addition, Groundwater Protection Standards will be established for all detected Appendix IV constituents. Per §257.95(h)(1-3), these standards will either correspond to the Maximum Contaminant Level (MCL) values that are provided within 40 CFR 141.62 and 141.66, or the background upper prediction limit for those constituents that do not have an MCL. In cases where the background upper prediction limit is higher than the MCL, the upper prediction limit will serve as the Groundwater Protection Standard.

For purposes of comparison, if the concentrations of all Appendix III constituents and the detected Appendix IV constituents are shown to be at or below the background prediction limits for two consecutive sampling events, the CCR unit will return to detection monitoring. The return to detection monitoring must be documented in the facility's operating record per §257.105(h)(7), noticed to the State Director per §257.106(h)(5), and posted on the publicly accessible internet site per §257.107(h)(5). If concentrations of any of these same constituents are above the background prediction limits but below the Groundwater Protection Standard, assessment monitoring will continue.

However, should one or more Appendix IV constituents exhibit a statistically significant increase above the Groundwater Protection Standard, documentation of the specific Appendix IV constituent(s) must be placed in the facility's operating record per §257.105(h)(8), along with notification to the State Director per §257.106(h)(6) and posting to the publicly accessible internet site per §257.107(h)(6). In addition, an investigation must be undertaken to evaluate the nature and extent of a possible release from the CCR unit and account for any other conditions that may factor into potential remedy implementation in accordance with the elements in §257.95(g)(1)(i-iv). Pending the outcome of the investigation, it may be possible to nullify the findings or identify an alternate source (similar to the process provided for under detection monitoring), and return the CCR unit directly to assessment monitoring. In the event that the CCR unit is ultimately deemed responsible for statistical increases in the groundwater constituent concentrations, the provisions of §257.96, §257.97, and §257.98 of the Rule would be followed to guide potential remedy assessment, selection and implementation. Moreover, should the responsible CCR unit be identified as an unlined surface impoundment, then the requirements under §257.95(g)(5), §257.101(a)(1), and §257.102 would be triggered for retrofit or closure.

4.0 Professional Engineer's Certification

In accordance with §257.93(f)(6) of the Rule, I hereby certify based on a review of the information contained herein, and my knowledge and understanding of the principles and accepted practices contained in EPA's Unified Guidance, that the statistical method selected for evaluation of groundwater data associated with the Keystone CCR Units is adequate and appropriate. This method's application will provide the necessary means for determining compliance and potential identification of statistically significant increases in groundwater concentrations as related to the Detection Monitoring (§257.94) and Assessment Monitoring (§257.95) elements of the Rule.

Certified by: _____

Date: 10/9/17

Richard Southorn, P.E., P.G., CPSWQ

Professional Engineer Registration No. PE 085411

CB&I Environmental & Infrastructure, Inc.

5.0 References

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

Helsel, D. R. and R. M. Hirsch, "Discussion of Applicability of the t-test for Detecting Trends in Water Quality Variables," Water Resources Bulletin 24, pp. 201-204, 1992.

United States Environmental Protection Agency, "Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities—Unified Guidance (EPA 530/R-09-007)," March 2009.

Figures

OFFICE	DATE	DESIGNED BY	DRAWN BY	CHECKED BY	APPROVED BY	DRAWING NUMBER
Pittsburgh, PA	*6/26/17*	--	*E. Schlegel*	--	--	**1009174011-B3**

LEGEND:

MW-5 CCR GROUNDWATER MONITORING WELL

SCALE

0 100 200 FEET

REFERENCES:

1. GOOGLE AERIAL PHOTOGRAPH, DATED 10/11/2015.

APTIM
500 Penn Center Boulevard,
Suite 1000
Pittsburgh, Pennsylvania 15235

GenOn™

FIGURE 1
CCR COMPLIANCE GROUNDWATER MONITORING WELL LOCATION MAP
ASH FILTER PONDS

KEYSTONE GENERATING STATION
PLUMCREEK TOWNSHIP, ARMSTRONG COUNTY, PA

File: O:\PROJECT\1009174011_keystone\1009174011-B3.dwg
Plot Date/Time: Sep 18, 2017 - 6:37am
Plotted By: greg.jones
Xref: Image

OFFICE	DATE	DESIGNED BY	DRAWN BY	CHECKED BY	APPROVED BY	DRAWING NUMBER
Pittsburgh, PA	6/26/17	--	E. Schlegel	--	--	1009174011-B4

LEGEND:

MP-18 CCR GROUNDWATER MONITORING WELL

SCALE

0 600 1,200 FEET

REFERENCES:

1. GOOGLE AERIAL PHOTOGRAPH, DATED 10/11/2015.

APTIM
500 Penn Center Boulevard,
Suite 1000
Pittsburgh, Pennsylvania 15235

GenOn™

FIGURE 2
CCR COMPLIANCE GROUNDWATER MONITORING WELL LOCATION MAP
EAST VALLEY AND WEST VALLEY ASH DISPOSAL SITES
KEYSTONE GENERATING STATION
PLUMCREEK TOWNSHIP, ARMSTRONG COUNTY, PA

File: O:\PROJECT\1009174011_keystone\1009174011-B4.dwg
Plot Date/Time: Sep 18, 2017 - 6:39am
Plotted By: greg.jones
Xref: Image