Groundwater Statistical Methods Certification

Neal South CCR Monofill
Permit No. 97-SDP-13-98P
Salix, Iowa

MidAmerican Energy Company

GHD | 11228 Aurora Avenue Des Moines Iowa 50322-7905

11114654 **|** Report No 5 **|** October 17 2017

Table of Contents

1. Introduction

Regular groundwater monitoring, sampling, and analysis is conducted at wells associated with the Neal South Energy Center CCR Monofill (Neal South CCR Monofill). Under the federal coal combustion residual (CCR) rule (40 CFR 257) (CCR Rule), the data obtained are to be subjected to statistical evaluation to demonstrate compliance with monitoring goals. Specifically, requirements for groundwater monitoring and corrective action are presented in sections 257.90 through 257.98 of the CCR Rule.

This report identifies and certifies appropriate evaluation methods as informed by United States Environmental Protection Agency's (USEPA's) "Unified Guidance Document: Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities," March 2009, EPA 530/R-09-007.

2. Assessment Strategy

The CCR Rule allows for different approaches to be used in statistically assessing groundwater monitoring data. For the purpose of performing an appropriate assessment of the Neal South CCR Monofill groundwater quality on an ongoing basis, the following strategy will be employed:

i) It is anticipated substantial spatial variability in groundwater conditions exists between monitoring wells, including upgradient background and downgradient wells. This will be tested for utilizing inter-well comparisons of the baseline data. Specifically, statistical upper tolerance limits (UTLs) will be calculated for each monitoring constituent considering the upgradient background data, and the downgradient well data will be compared against these.

ii) If statistically significant spatial variability is demonstrated, ongoing assessment of future monitoring data will be conducted through intra-well comparisons, in which statistical UTLs will be calculated for each monitoring constituent at each well considering the baseline period data, against which future monitoring results from the same well will be compared.

This assessment requires a number of different methods and evaluation components, the specifics of which are described in the following report sections.

3. Assessment of Statistical Assumptions and Data Screening

As part of conducting intra-well and inter-well comparisons of monitoring data as described in Unified Guidance (USEPA, 2009), characteristics of each data set considered must be assessed in order to select appropriate statistical analysis methods. In particular, one must assess: (i) the observed data distribution, (ii) the percentage of censored data (non-detect results) present, and (iii) the presence of statistical outliers. These issues are discussed in Unified Guidance, and methods for assessing these characteristics are provided in USEPA's ProUCL software[1]. Additional information on the statistical methods available is presented in the ProUCL Technical Guide (USEPA, 2015), and a brief introduction is provided below.

ProUCL assesses each data set considered for the following distribution patterns (in priority order): normal; gamma; and lognormal. If a data set is found to be described by one of these distributions, then a statistical method appropriate for the observed data distribution is used. If, however, a particular data set does not follow one of these distributions, it is identified as having an unknown distribution, and non-parametric (rank-based) statistical methods are used for subsequent calculations.

Once a data distribution has been established for a data set, an assessment of statistical outliers (extreme low or high values appearing atypical of the remaining data) is carried out considering the observed data distribution. Initial screening of potential outliers can be conducted graphically, utilizing quantile/probability plots. Confirmation of suspected outliers may then be performed using formal statistical tests, such as Dixon's extreme value test (for less than 25 observations) or Rosner's test (for more than 25 observations). Details of these methods are found in Chapter 12 of Unified Guidance (USEPA, 2009) and Section 7.3 of the ProUCL Technical Guide (USEPA, 2015).

If censored data (non-detects) are present in a data set, suitable methods are required to perform an appropriate statistical analysis. For statistical procedures based on mean and standard deviation calculations, the Kaplan-Meier (KM) method is recommended (USEPA, 2015; Helsel, 2005), which may be used when single or multiple detection limits are present. The method is described in Section 5.3 of the ProUCL Technical Guide (USEPA, 2015). For statistical procedures requiring assignment of magnitudes to individual observations (e.g., trend analysis), data substitution (e.g., detection limit or fraction thereof) may be appropriate if very few non-detects are present (e.g., not more than 10-15 percent of the data set, per Section 15.6 of the Unified Guidance), but non-parametric tests based on ranking of the non-detect data (as less than detected values and tied with other non-detects) is preferable.

For heavily-censored data sets (i.e., those containing high proportions of non-detect results), statistical procedures for evaluating data distributions and outliers lose sensitivity, and estimating required parameters such as means and standard deviations becomes a speculative process. In such situations, the Unified Guidance and ProUCL Technical Guide recommend the use of non-parametric (rank-based) procedures. Non-parametric methods do not make distributional assumptions, and are not influenced by a small number of outlier values, if present.

3.1 Field Duplicates

Field duplicate results (replicates) do not represent statistically-independent results, and may skew the weighting of a sampling event where a field duplicate result was collected, as compared to an event where only a single investigative sample is taken. However, both field duplicate results do provide useful estimates of the constituent concentration present in groundwater during a given event. Therefore, field duplicate results should be averaged prior to completing statistical analyses. If one field duplicate is a detected value and the other a non-detect, it is recommended that the

[1] Current version 5.1 (5.1.002). Available at https://www.epa.gov/land-research/proucl-software

detected result be conservatively retained to represent a maximum estimate of the analyte concentration.

The specific statistical procedures employed are given below with the various assessment components.

4. Stability Assessment / Baseline Period Trend Analysis

The assessment of temporal trends is included as a precursor to intra-well and inter-well comparisons, as the statistical methods for these comparisons assume that a stable condition is present in the reference data set (i.e., the baseline period for intra-well comparisons, and the upgradient background data set for inter-well comparisons). Where a trend is identified, the comparison procedures must be adjusted to take this into account. Trend tests may also be used as an alternative intra-well comparison procedure when other methods are not appropriate (as noted at the beginning of Section 17.3 of the Unified Guidance (USEPA, 2009).

Statistical procedures for evaluating trends are discussed in USEPA and United States Geological Survey (USGS) guidance documents (USEPA, 2006, 2009; USGS, 2002) and a variety of environmental statistic texts. Different trend test procedures are available for a variety of purposes and data set characteristics. The selection of an appropriate test to apply depends on satisfying any underlying assumptions of the statistical method, as well as the type of hypothesized trend investigated.

It is anticipated that two primary statistical trend test procedures will be used in the stability/trend assessment of the baseline monitoring data, based on characteristics of each data set. These include:

- The Mann-Kendall trend test – for data sets with 0-50 percent non-detects, to determine if a statistically significant increasing or decreasing trend is evident during the baseline period
- The Theil-Sen median slope – for those data sets with statistically significant trends identified by the Mann-Kendall test, a slope estimate will be provided of the magnitude (rate) of the trend

The Mann-Kendall and Thiel Sen procedures can accommodate moderate levels of censored data (e.g., up to 50 percent non-detects).

The anticipated procedures are described below.

4.1 Mann-Kendall Trend Test

The Mann-Kendall test, which is commonly applied to environmental monitoring data (USGS, 2002; USEPA, 2009) is a non-parametric (rank-based) method that evaluates a set of data for a monotonic (unidirectional) trend result. The procedure makes no assumptions regarding the shape of the trend (e.g., linear, log linear), except that the trend is in a single direction (i.e., either consistently upward or downward).

Directions for carrying out the Mann-Kendall test are provided in Section 17.3.2 of the Unified Guidance (USEPA, 2009). In general terms, the test compares each possible pair of data points (e.g., if there are four observations through time, then 3+2+1=6 pairwise comparisons are made), recording whether a concentration went up ("+") or down ("-") between the earlier and later data points. The test then sums the results (e.g., 4 "+" and 2 "-" yields a net of +2) and compares this value to a standard reference table to determine the statistical significance based on the number of observations (e.g., Table 17-5 of USEPA, 2009).

4.2 Theil-Sen Slope Estimate

The Theil-Sen slope procedure is a non-parametric alternative to linear regression that may be used in conjunction with the Mann-Kendall test, in order to obtain an estimate of the magnitude of statistically significant trends (USEPA, 2009). Like the Mann-Kendall test, the Theil-Sen trend line is a non-parametric procedure that employs rank-based statistics. As such, the procedure is robust even in the presence of statistical outliers and moderate levels of censored (non-detect) data.

Directions for calculating the Theil-Sen trend line are provided in Section 17.3.3 of the Unified Guidance (USEPA, 2009). The procedure involves computing slopes (the change in concentration divided by the change in time) for each possible pair of data points, and then selecting the median slope as the overall estimate. A Theil-Sen trend line may be constructed if desired using the median slope to establish a line through a point located at the median concentration and median time of the data set being assessed.

4.3 Statistical Significance Levels

In carrying out trend tests, a target significance level of 0.05 (i.e., 95 percent confidence) will be used. This significance level is applied on a per-location, per-parameter basis, for all suitable data sets. The CCR Rule (40 CFR 257.93 (g)(2)) specifies that a significance level no lower than 0.01 (99 percent confidence) may be applied for individual comparisons, with an overall (site-wide) significance level of 0.05. This requirement is not specifically specified for trend testing, but is useful as a frame of reference for selecting a suitable significance level. By selecting a significance level of 0.05, the statistical power of detecting potential trends with data sets containing only eight baseline samples (which is the number of baseline events specified in the CCR Rule) will increase (as compared to using a significance of 0.01).

4.4 Consideration of Censored (Non-Detect) Data

As noted previously, both the Mann-Kendall and Thiel-Sen procedures can accommodate moderate levels of censored data (e.g., up to 50 percent non-detects). Since these tests are focused on trends related to the median values of a given dataset, the tests have reasonable sensitivity as long as the central values (medians) are based on detected concentrations.

The appropriate treatment of censored data varies slightly between the trend procedures employed:

1. For the Mann-Kendall trend test, non-detects are considered to be tied (i.e., equal) values with concentrations lower than the detected concentrations. For convenience, a value of zero may be used for the non-detects, although any value below the lowest detected result would yield

identical ranking in the test (as a non-parametric method, Mann-Kendall considers only whether a certain observation is above or below another, and not the magnitude of the difference).

2. For the Theil-Sen trend line determination, non-detects will be assigned a value equal to the reported detection limit. This is a conservative approach, in that the magnitude of slope estimates involving non-detects represents a lower bound on the possible slope (e.g., compared to using zero or a value between zero and the detection limit). As noted above, the Theil-Sen trend line utilizes the median slope estimate, which is based on detected values for the vast majority of data sets containing up to 50 percent non-detects. Thus, the substitution of detection limits for non-detect values has minimal or no impact on the Theil-Sen trend line in the present analysis.

In some cases, data points may be present that have ambiguous rankings compared to other data points. Typically, this occurs when there is either:

- A low detected value (typically "J-qualified") below the detection limits of non-detects, or
- An elevated detection limit (e.g., due to matrix interference) above a number of detected values

When a data set contains estimated data (J-qualified data) and/or non-detects (NDs) with elevated detection limits [e.g., ND (1.0) vs. 0.60 J], the comparison of NDs with higher detection limits to a lower detected value can lead to an erroneous calculation of a trend. The approach used by the ProUCL Technical Guide (USEPA, 2015) is to censor all low data to the highest detection limit [e.g., 0.60 J is treated as ND (1.0)]. However, this may result in many detects being censored, in which case a trend test may not be possible.

The recommended approach for cases where ambiguous rankings exist in a dataset is as follows:

- If one or a few low estimated (J-qualified) detected values are present below detection limits, these should be considered to be non-detects, as long as the resulting data set does not contain more than 50 percent non-detects.
- Where one or a few non-detects with elevated detection limits are present above detected values, these should be excluded from the evaluation, as long as the resulting data set consists of at least four observations, and at least two of the detected values are not J-qualified (i.e., estimated values).
- Where moderate numbers of low detected values and non-detects with higher detection limits are present, such that censoring these to a common detection limit would result in an untestable dataset (all data considered as non-detects), testing should be performed considering the detects only, as long as a minimum of four observations are retained including at least two detected values that are not J-qualified (i.e., estimated).

If ambiguous rankings may not be resolved in one of these ways, trend testing would not be performed.

5. Statistical Comparison Procedures

Included in the CCR Rule (specifically, 40 CFR 257.93(f)), and more fully explained in the Unified Guidance (USEPA, 2009), statistical comparisons may be conducted via one or more statistical methods listed therein, including:

- Parametric ANOVA
- Non-parametric (rank-based) ANOVA
- Tolerance or Prediction Limits
- Statistical Control Charts
- Another suitable test meeting the performance requirements

Each of these procedures has different requirements and applicability for monitoring data assessment. For the purposes evaluating the Neal South CCR Monofill groundwater monitoring data, a tolerance limit approach will be used.

A discussion of the derivation and application of statistical tolerance limits for evaluating groundwater monitoring data is found in Section 17.2 of Unified Guidance (USEPA, 2009), with further relevant information present in Chapter 5. Note that the statistical procedures (tolerance limits) may be used for both inter-well and intra-well comparisons (where appropriate, e.g., due to demonstrated spatial variability). For intra-well comparisons, the upgradient background data are used as a reference group. For inter-well comparisons, the baseline period is used as a reference group against which future data are compared. To distinguish between these two cases herein, 'upgradient background' is used when referring to inter-well comparisons, and 'baseline' is used when referring to intra-well comparisons.

A statistical tolerance limit, and specifically an upper tolerance limit (UTL), is a statistically-based limit above which a given sample measurement is unlikely to occur if conditions are consistent with the reference population. The general approach found in the Unified Guidance and the ProUCL Technical Guide (USEPA, 2009, 2015) for determining UTLs is to estimate an upper bound on the reference population using a method appropriate for the observed data distribution (i.e., normal, gamma-distributed, lognormal, or none of these) and degree of censoring (non-detects) present. UTLs take into consideration sampling variability (both in reference and comparison group samples), and provide values which future samples have a low probability of exceeding by random chance (e.g., less than 5 percent if a 95 percent coverage is used) if conditions remain consistent with the baseline or upgradient background reference group.

Statistical UTLs have two descriptive parameters: (i) their *coverage*; and (ii) their *confidence*. Coverage refers to the percentage of the reference population that is within the tolerance limit. For example, if a coverage of 0.99 is selected, then 99 percent of the reference population will be within the tolerance limit (which in the case of a one-sided UTL, represents the 99th percentile of the reference population). Confidence refers to the probability that the specified coverage based on the samples collected will include the true value from the entire population. For example, if a 95 percent confidence level is selected, then there is no more than a 5 percent probability that the calculated UTL will not include the selected coverage (e.g., true 99th percentile) of the population.

For the purposes of evaluating the groundwater monitoring data, two UTLs are considered:

i) A 95/95 UTL, which has a coverage of 0.95 (i.e., 95th percentile) with 95 percent confidence

ii) A 99/95 UTL, which has a coverage of 0.99 (i.e., 99th percentile) with 95 percent confidence

The utility of considering two UTLs is in the response to future observations above a UTL. Per the Unified Guidance (USEPA, 2009), it is customary to collect a verification resample for confirmation of any monitoring data that exceed a UTL before identifying an SSI. That is, a single occurrence of a future value above a UTL is not considered to represent an SSI unless confirmed by one of more verification resamples (see Section 4.3 and Chapter 19 of the Unified Guidance [USEPA, 2009], discussing this issue in the context of upper prediction limits and is equally applicable to UTLs).

Since a 95/95 UTL has up to a 5 percent chance of observing a future sample above it due to natural variability, and that multiple comparisons (i.e., many constituents and wells) are conducted for each monitoring event, it is not unlikely that one or more observations may occur above a 95/95 UTL without a significant underlying concentration increase. There is much less chance (only up to one percent) that a 99/95 UTL will be exceeded without an underlying change in conditions. Thus, the following evaluation approach and response is proposed:

a) If a given future observation exceeds a 95/95 UTL, but not a 99/95 UTL, the next regularly scheduled monitoring event will be used as the verification resample for determining an SSI.

b) If a given future observation exceeds both the 95/95 and 99/95 UTLs, then a verification resample or other appropriate response will be undertaken prior to the next regularly scheduled monitoring event in order to make a quicker determination if an SSI has occurred.

If two subsequent samples (regular monitoring events and/or special verification resample events) are observed to have groundwater containing a constituent above a 95/95 UTL, then an SSI increase will be identified and applicable response required under the CCR Rule initiated within 90 days, per 40 CFR 257.94(e).

The calculation of tolerance limits is presented in detail in Chapter 3 (for data sets without non-detects) and Chapter 5 (for data sets with non-detects) of the USEPA's ProUCL Technical Guide (USEPA, 2015). The present evaluation utilizes the logic and methods provided in ProUCL (version 5.1.002).

Where temporal trends are identified in a data set over the baseline period, tolerance limits may not be calculated, as they assume a stationary population. In such cases, the baseline range may be utilized as a reference. For such data sets, future sample results should be compared both against the baseline range and what would be expected based on the observed trend over the baseline period.

5.1 Testing for Spatial Variability – Inter-well Comparisons

One of the purposes of conducting baseline monitoring is to identify existing spatial differences in groundwater conditions between wells. The first strategy for identifying SSIs under the CCR Rule is through inter-well comparisons, in which data at downgradient well locations are compared against upgradient background. However, where spatial variability is present it is necessary to employ intra-well comparisons, in which data at each well location are compared against baseline reference conditions at that well.

To test for spatial variability in the groundwater monitoring data collected at the site, inter-well background UTLs will be calculated pooling data from the two upgradient wells at the site (MW-4 and MW-15). Upgradient background UTLs will be calculated following the procedures described above, based on the Unified Guidance (USEPA, 2009) and ProUCL software (USEPA, 2015) methods.

The calculated upgradient background UTLs will be summarized, and then the data from downgradient wells will be compared against these on a point-by-point basis. Where a downgradient well has constituent concentrations consistently above the upgradient background UTLs (i.e., in two or more subsequent sample during the baseline period), a significant spatial difference will be noted. Future sample data may similarly be compared to the upgradient background UTLs to look for SSIs over inter-well background conditions.

5.2 Accounting for Spatial Variability – Intra-well Comparisons

Where spatial variability is demonstrated between downgradient and upgradient well locations over the baseline period, this needs to be accounted for in the data assessment. Specifically, intra-well comparisons are appropriate when significant spatial variability is present.

In order to provide a framework for intra-well comparisons, baseline reference values (95/95 and 99/95 UTLs) will be calculated on a per-well, per-constituent basis. The methodology used will be the same as for inter-well upgradient background values, following the procedures in the Unified Guidance (USEPA, 2009) and ProUCL software (USEPA, 2015). These calculations will provide a basis for assessing future data (post-baseline) to identify SSIs over intra-well baseline conditions. As noted above, if two subsequent samples are observed exceeding calculated baseline UTLs, a significant increase will be noted.

For both intra-well and inter-well comparisons, if a significant trend over time is detected over the baseline period, then assessment of future data will be conducted through comparisons to the baseline range and ongoing trend testing.

6. Alternative Statistical Methods

The proposed statistical methods described above have been selected for the observed baseline data and are anticipated to cover the variety of data obtained (different data distributions, non-detect data presence, etc.). However, if future changes to conditions are observed, for example regional trends in constituent concentrations unrelated to the Neal South CCR Monofill, MidAmerican reserves the ability to adjust the statistical methodology employed to address such changes and update this certification. In such cases, the guiding principles for selecting appropriate statistical evaluation methods will be:

a) Consistency with the CCR Rule

b) Consistency with USEPA's Unified Guidance (2009) or updates thereto

c) Applicability and appropriateness to the data distributions and characteristics observed for each well and constituent monitored

7. Certification

I certify the selected statistical methods are appropriate for evaluating the groundwater monitoring data for the Neal South CCR Monofill.

LICENSED PROFESSIONAL ENGINEER MICHAEL J. ALOWITZ 18160 IOWA	**I hereby certify that this engineering document was prepared by me or under my direct personal supervision and that I am a duly licensed Professional Engineer under the laws of the State of Iowa.** **Michael J. Alowitz, P.E.** — **Date:** 10/17/17 **License Number:** 18160 **My license renewal date is:** December 31, 2018 **Pages or sheets covered by this seal:** Entire Document

8. References

Helsel, D. R., 2005. Nondetects and Data Analysis. Statistics for Censored Environmental Data. John Wiley and Sons, N.Y.

USEPA, February 2006. Data Quality Assessment: Statistical Methods for Practitioners (EPA QA/G-9S). Office of Environmental Information, United States Environmental Protection Agency, Washington D.C. EPA/240/B-06/003.

USEPA, March 2009. Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities – Unified Guidance. Office of Resource Conservation and Recovery, Program Implementation and Information Division, United States Environmental Protection Agency, Washington D.C. EPA/530/R-09/007.

USEPA, October 2015. ProUCL Version 5.1.002 Technical Guide. United States Environmental Protection Agency, Office of Research and Development, Washington D.C. EPA/600/R-07/041.

USGS, 2002. Statistical Methods in Water Resources. By D.R. Helsel and R.M. Hirsch. Chapter A3 of Book 4, Hydrologic Analysis and Interpretation in Techniques of Water Resources Investigations of the United States Geological Survey.

www.ghd.com