Statistical Method Selection Certification

R.M. Heskett Station

Prepared for
Montana-Dakota Utilities Co.

October 2017

Statistical Method Selection Certification

R.M. Heskett Station

Prepared for
Montana-Dakota Utilities Co.

October 2017

4300 MarketPointe Drive, Suite 200
Minneapolis, MN 55435
952.832.2600
www.barr.com

Statistical Method Selection Certification

R.M. Heskett Station

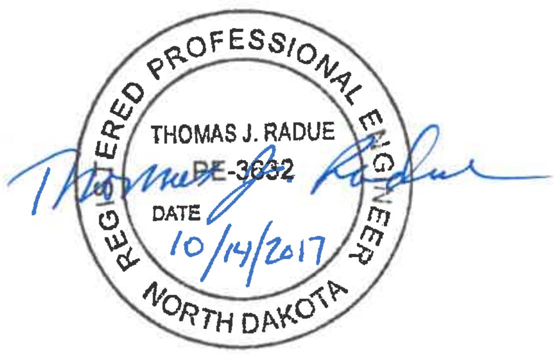
October 2017

Table of Contents

Certifications

I hereby certify that I have examined the facility and, being familiar with the provisions of 40 CFR 257 Subpart D, attest that this Statistical Method Selection Certification has been prepared in accordance with good engineering practice, including consideration of applicable industry standards and the requirements of 40 CFR §257.93. I hereby certify that the plan certified herein is adequate for this facility and that procedures for recordkeeping and reporting have been established. I further certify that I am a duly Licensed Professional Engineer under the laws of the state of North Dakota.

Thomas J. Radue, P.E.
PE #: 3632



Revision	Date	Summary of Revisions
0	October 14, 2017	Initial Groundwater Statistical Analysis Plan

Acronyms

Acronym	Description
ANOVA	analysis of variance
CCR	coal combustion residual
KM	Kaplan-Meier
ND	non-detect
QA	quality assurance
QC	quality control
SAP	samplings and analysis plan
SSI	statistically significant increase

1.0 Introduction

This Statistical Method Selection Certification (Certification) describes part of the Sampling and Analysis program (Program) required by the CCR Rule (EPA, 2015a) §257.93 Groundwater Sampling and Analysis Requirements. The Certification provides a description of the potential statistical methods that may be used to evaluate groundwater quality data collected from the monitoring system for the CCR unit at the R.M. Heskett Station (Site) located in Mandan, North Dakota.

Statistical evaluations of groundwater samples will be conducted in accordance with CCR Rule Section 257.93 (f) as well as the preamble, which states that "Guidance on selecting a specific method is described in "Unified Guidance Document: Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities," March 2009, EPA 530/R-09-007." (pg. 21402). Additional technical guidance documents utilized to form decisions on the statistical evaluation methodology include:

- ITRC, 2013. ITRC Guidance Document: Groundwater Statistics and Monitoring Compliance.
- EPA, 2015. ProUCL Version 5.1.002 Technical Guide: Statistical Software for Environmental Applications for Data Sets with and without Nondetect Observations. EPA/600/R-07/041. October 2015.
- ASTM, 2017. D6312-17 Standard Guide for Developing Appropriate Statistical Approaches for Groundwater Detection Monitoring Programs at Waste Disposal Facilities, ASTM International, West Conshohocken, PA, www.astm.org.

Background groundwater quality concentrations will be utilized to determine if there is a statistically significant increase (SSI) in downgradient wells for Appendix III, and if necessary Appendix IV, parameters listed in the CCR Rule. Depending on Site conditions, methods such as interwell and/or intrawell evaluations may be utilized to evaluate for an SSI for each parameter.

In general, the most appropriate methods for evaluating groundwater data involve comparison to prediction limits or control charts (ASTM, 2017; EPA, 2009), however specific conditions may warrant the utilization of the other methods included in the CCR Rule Section 257.93 (f). The consideration of which types of prediction limits (e.g. interwell vs. intrawell) or control charts will be based on evaluation of the distributional characteristics of the baseline data, the number of non-detects, spatial variability, trends, outliers, etc. As allowed by the CCR Rule Section 257.93 (g)(1), statistical methods may include more than one type of analysis depending on the data distribution for each well/parameter combination.

The purpose of this Statistical Method Selection Certification is to:

1. Provide a means to consistently evaluate the data and to establish criteria that will identify SSIs that may indicate a potential release from the facility; and
2. Define a protocol for verification of SSIs.

2.0 Data Evaluation Methods

2.1 Preliminary Data Evaluation

Preliminary data evaluations may be necessary to evaluate background data and in advance of or to supplement the SSI evaluation. The first consideration is to review or modify the existing Site Conceptual Groundwater Flow Model to assess whether it provides an overall fit to the laboratory data and to develop preliminary data hypotheses. The Site Conceptual Groundwater Flow Model provides a framework to understand how a hypothetical release to groundwater from the CCR unit would flow through the groundwater monitoring system. Next the data for each event will be evaluated using standard hydrogeologic methods to consider overall flow directions, well depths, locations, and likely flow paths relative to the monitored CCR unit.

2.1.1 Data Summary Statistics

The summary statistics may include number of observations, number of values below the detection limit (non-detects), minimums, maximums, means, medians, measures of variability, and quantiles. Graphical displays, such as box plots and Q-Q plots, may be utilized to visually evaluate the data, including the presence and degree of the outlier(s). In box plots, the box represents 25^{th} and 75^{th} percentiles, the middle line shows the median (50^{th} percentile), the whiskers represent the entire range of the data.

2.1.2 Data Distribution

It is necessary to define the type of distribution (normal, log-normal, non-parametric, etc.) for each parameter at each well prior to choosing statistical analysis testing methods. Testing methods may be more or less appropriate for different types of distributions. A Goodness-of-Fit test will be performed at a significance level of 0.05. Data distribution determination will be updated each time statistical analyses are conducted.

2.1.3 Outlier Testing

Outlier testing can be conducted on data sets that are normally distributed. Outliers will be evaluated by the Dixon and Rosner test methods, both of which are outlined in the EPA Unified Guidance (EPA, 2009). Outliers will not be evaluated statistically for non-normally distributed data sets.

2.1.4 Charge Balance Error Calculation

Charge balance error calculations are utilized to determine if the cations and anions are balanced or electrically neutral. To do this, the concentrations of the major cations and anions (sodium, potassium, calcium, magnesium, alkalinity as bicarbonate, chloride, sulfate, and fluoride) in milligrams per liter (mg/L) are converted to milliequivalents per liter (meq/L). This is done by multiplying the concentration of each parameter in mg/L by the electronic charge of the ion divided by the molecular weight of the cation or anion. Then the total meq/L of the cations and anions are summed and a percent difference is calculated. Ideally, the charge balance error between the cations and anions should be within ±15%. If a greater error is found, then there may be an issue with the laboratory analytical procedures and the owner/operator may wish to consult with the laboratory. Charge balance error calculation will be performed, dependent on availability of concentration data for major cations and anions.

2.2 Statistical Tests

Following the collection and analysis of the groundwater samples, Appendix III and/or Appendix IV parameters will be analyzed using statistical methods to determine if there are SSIs in each parameter concentration for each downgradient well relative to either the background (pooled upgradient wells; interwell) or baseline (historical data collected at the individual well; intrawell). The proposed statistical methods will be in accordance with the guidance documents listed in Section 1.0.

The CCR Rule allows for several types of statistical evaluations to be conducted to determine if an SSI has occurred for individual parameters in each downgradient monitoring well. This includes the following tests:

1. Parametric analysis of variance (ANOVA) followed by multiple comparison procedures (257.93 (f)(1));
2. Analysis of variance based on ranks followed by multiple comparison procedures (257.93 (f)(2));
3. Tolerance or prediction interval (257.93 (f)(3));
4. Control chart (257.93 (f)(4)); and
5. Another statistical test method that meets the performance standards outlined in 257.93 (g).

The performance standards of the statistical method as outlined in Section 257.93 (g) must:

1. Be appropriate for the distribution of the data;
2. Minimize the Type I (false positive) error: if an individual well comparison method is used, the Type I error level must not exceed 0.01 and if a multiple comparison method is used, the Type I error level must not exceed 0.05;
3. Control charts, tolerance intervals, and prediction intervals, if used, must be at least as effective as any other approach;
4. Account for non-detect data, utilizing the lowest concentration reasonably achieved within the specified limits of precision; and
5. Control and/or correct seasonal and temporal correlations.

Additional statistical test methods to evaluate an SSI may also include trend tests. Control charts and prediction limits can be employed as interwell and intrawell tests; that is, they can evaluate changes between well groups (such as upgradient vs. downgradient) or within a single well (historical vs. current conditions). Different conditions prompt the utilization of interwell or intrawell testing. A narrative description of the various statistical testing methods are outlined below.

The CCR Rule states that "The owner or operator of the CCR unit must select one of the statistical methods specified in paragraphs (f)(1) through (5) of this section to be used in evaluating groundwater monitoring data for each specified constituent. The statistical test chosen shall be conducted separately for each constituent in each monitoring well" (257.93 (f)).

The most appropriate statistical method for each constituent in each monitoring well will be determined each time a statistical evaluation is conducted. Additionally, because conditions may change over time, the method best suited for the individual constituents may change with time as well. Therefore, below is a discussion of the statistical methods allowed per the CCR Rule in Section 257.93 (f).

2.2.1 Interwell and Intrawell

Selection of the appropriate method (interwell vs. intrawell) depends on site-specific conditions. SSI evaluations for an individual parameter can be conducted using either of two methods: interwell analysis (comparison of pooled upgradient to individual downgradient wells) or intrawell analysis (comparison of baseline/background data to compliance data within a single well). Geologic environments are naturally heterogeneous in that differences in lithology exist at a variety of scales in the subsurface. Therefore, assumptions that the geology below a site will be homogenous and isotropic are not generally an accurate description of Site geology.

This observation is made more apparent because the components of leachate from CCR units all occur naturally in the subsurface and can be assumed to be present naturally at varying concentrations for all CCR Rule Appendix III and Appendix IV parameters. In some cases the natural differences aggregate into a relatively consistent geochemical signature. That signature can change if there is a change in chemical equilibrium (e.g. either natural process or a release from a CCR unit).

The Unified Guidance (EPA, 2009) contains a detailed discussion of tests and considerations for selection of the appropriate method that advocate evaluation of the site conceptual model and consideration of the geochemical conditions in each well and between up- and down-gradient wells. The general procedure for method selection will be to evaluate interwell methods first for fit to statistical assumptions and then identify those well/parameter combinations that appear better suited for intrawell methods. For efficiency, the preference will be to settle on one set of tests on all of the data, however there may be instances where a mixture of inter- and intra-well methods will be used to address unique characteristics of individual well/parameter combinations. The discussion below describes some of the tools available in the CCR Rule.

2.2.2 Analysis of Variance (ANOVA)

ANOVA is a test used to determine if there is a statistically significant difference between the mean concentrations of specified wells or groups of wells. ANOVA is included in the CCR Rule as one of the acceptable methods for evaluating SSIs, however the Unified Guidance states that it does not recommend this method for “formally making regulatory decisions about compliance wells or regulated units”; rather, the Guidance recommends utilization of prediction limits, control charts, and confidence intervals (discussed in sections below; EPA 2009, pg. 17-1). However, ANOVA has been identified as a useful tool to identify spatial variation.

ANOVA tests can be conducted on normal and non-normally distributed populations. The One-Way Parametric F-Test is most appropriate for normally distributed datasets whereas the Kruskal-Wallis Test is for non-parametric datasets. Both tests can be utilized to evaluate spatial variations.

2.2.3 Tolerance Intervals

A tolerance interval is defined as the "concentration range designed to contain a pre-specified proportion of the underlying population from which the statistical sample is drawn (e.g. 95 percent of all possible population measurements)" (EPA 2009, pg. 17-14). Two values are reported with the interval; the coverage (proportion of the population covered by the interval) and the degree of confidence or confidence level.

However, as with the ANOVA test method, the Unified Guidance does not recommend utilization of tolerance limits in making regulatory decisions.

2.2.4 Prediction Intervals

Prediction intervals (or limits) are utilized to predict the "upper limit of possible future values based on a background or baseline data set and comparing that limit to compliance point measurements or statistics…prediction 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" (ITRC 2013, pg. 87). Both parametric and non-parametric prediction limits can be determined. For parametric prediction limits, the mean and standard deviation of the background or baseline dataset is used whereas non-parametric prediction limits are based on order statistics (i.e. highest detected concentration; ITRC 2013, pg. 87). It should be noted that non-parametric prediction limits require a larger sample size than parametric prediction limits in order to obtain a desired confidence level. Additionally, for intrawell comparisons, ASTM recommends at least 13 baseline samples prior to reporting the nonparametric limit or utilize a Poisson Prediction Limit (ASTM, 2017). The Unified Guidance, however, states that the intrawell approach can be utilized "despite its inadequate power, until the intrawell background dataset is sufficiently large via periodic updates" (EPA 2009, pg. 6-34). Finally, inclusion of outliers should be avoided, either through removal from the dataset or, for non-parametric limits, utilize a different order statistic, such as the second highest detected concentration (EPA 2009, pg. 18-17).

The ASTM Standard Guide suggests the utilization of prediction limits when conducting interwell comparisons and for nonparametric intrawell comparisons. The appropriate methodology of calculating a prediction limit is dependent on detection frequency and data distribution. For datasets containing more than 50% detections either normal, log-normal, or non-parametric prediction limits should be utilized, depending on the data distribution. For datasets containing less than 50% detections, a non-parametric prediction limit should be utilized. For datasets containing no detections, the prediction limit should equal the detection limit. The ASTM Standard Guide also recommends computing the false positive and negative rates following the calculation of the limit and if the false positive rate is greater than 5%, several changes are advised, including increasing the number of background samples, changing the verification resampling plan, etc. (ASTM, 2017).

2.2.5 Control Charts

Control charts are constructed similar to prediction limits, where the limit is estimated from background or baseline and compared to compliance measurements. They are designed for datasets with known distributions and a high detection frequency (>85% detection; EPA 2009, pg. 20-10). Although not specified in the CCR Rule, the Unified Guidance recommends the utilization of the Shewhart-CUSUM control chart (EPA 2009, pg. 6-46). The Shewhart-CUSUM control chart conducts two tests against the control limit, the Shewhart control line, which identifies sudden spikes in concentrations or changes in trend (such as a new release), and the cumulative sum control chart (CUSUM), which identifies more gradual trend changes over time (plume migration; ITRC 2013, pg. 136). The utilization of the Shewhart-CUSUM control chart (vs. a non-Shewhart-CUSUM control chart) will be evaluated at the time of analysis.

2.2.6 Verification Retesting

Based on the CCR Rule timeline and guidance, the 1-of-2 retesting strategy may be utilized, or another appropriate retesting strategy supported by guidance. That is, if a limit exceedance is observed, a verification resample will be collected and compared to the limit. If the limit exceedance is verified, then an SSI will be declared. However, if the exceedance is not verified, the initial sample will be flagged and not included in subsequent analysis and the verification resample will be utilized in its place (EPA 2009, pg. 20-15).

2.3 Data Considerations

2.3.1 Non-Detects

Parameters with concentrations below the detection limit (non-detect data) will be reported at the detection limit and will be appropriately accounted for during the statistical analysis. No substitution method will be utilized in the handling of non-detect data.

2.3.2 Duplicates

Duplicate samples are typically collected as a quality control check on field and laboratory sampling methods, not for data analysis purposes. If the owner/operator conducts a QA/QC review of the data and the results of a duplicate sample do not sufficiently match that of the original sample, a qualifier is typically applied and, based on the type of qualifier, the original data may be excluded from further analysis. Duplicate sample results will not be utilized in data evaluations beyond QA/QC procedures.

2.3.3 Seasonal Trends

As required by the CCR Rule, seasonal trends will be evaluated and properly adjusted for, if detected.

3.0 SSI Evaluation

For the purpose of this Site, some combination of control charts and prediction limits will be used to determine if an SSI has occurred, along with other statistical evaluation tools such as ANOVA, box plots, and/or trend tests. Determination of an SSI in downgradient monitoring wells will be based on whether or not the control chart or prediction limit for a given parameter has been exceeded at a downgradient well. Because the control charts and prediction limits are based on a minimum of eight events, and comparison of the ninth sample to the limits established by the preceding eight events, no SSIs can be evaluated until after the ninth sample has been collected and analyzed.

3.1 Data Validation

Samples are sent to the laboratory for analysis immediately after collection. Therefore, the first decision in SSI evaluation is to determine that the data provided by the laboratory is representative of the groundwater. For this reason a series of checks will be conducted to assess data quality prior to performing statistical evaluations.

Following the receipt of the final analytical report from the laboratory for all the monitoring locations and QA/QC evaluation, a preliminary evaluation of the data will be conducted to ensure no issues with the analytical data prior to further analysis. This may include calculating the charge balance error, comparison of total and dissolved concentrations, and/or outlier analysis.

3.2 Profile Statistics

Once the validity of the data has been verified, time series plots, summary statistics, stiff and piper plots, a distribution analysis, and/or a trend test may be conducted. If a trend is detected, additional evaluations may be warranted to determine if there is an off-site/non-CCR unit causing contamination.

3.3 Statistical Testing

Depending on Site conditions and spatial variability of the parameters, inter- and/or intra-well testing procedures will be utilized to compare compliance data to the appropriate limits. As previously stated, spatial variability for the individual parameters may be evaluated using either box and whisker plots and/or ANOVA testing. The ASTM standard recommends utilizing prediction limits when interwell analysis is conducted and a combination of control charts and prediction limits when intrawell analysis is conducted (ASTM, 2017).

If an exceedance of a limit occurs in one or more downgradient wells, the well(s) where the suspected exceedance occurred will be resampled. As previously stated, the 1-of-2 retesting strategy may be utilized, or another appropriate retesting strategy supported by guidance. As recommended by ASTM, "a statistically significant exceedance is not declared and should not be reported until the results of the verification resample are known" (ASTM 2017, pg. 9). If the verification resample for a given parameter at a given well exceeds the control limit, then an SSI will be declared for that well-parameter pair. If not, detection monitoring will continue.

4.0 References

ASTM, 2017. D6312-17 Standard Guide for Developing Appropriate Statistical Approaches for Groundwater Detection Monitoring Programs at Waste Disposal Facilities, ASTM International, West Conshohocken, PA, www.astm.org.

Hensel, D.R. and R. M. Hirsch, 2002. Statistical Methods in Water Resources Techniques of Water Resources Investigations, Book 4, chapter A3. U.S. Geological Survey. 522 pages.

ITRC, 2013. ITRC Guidance Document: Groundwater Statistics and Monitoring Compliance.

EPA, 2009. Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities: Unified Guidance. EPA 530/R-09-007. March 2009.

US EPA, 2010. National Functional Guidelines for Inorganic Superfund Data Review, US EPA Contract Laboratory Program, United States Environmental Protection Agency. [Online] OSWER 9240.1-51. USEPA-540-R-10-011, January 2010. http://epa.gov/superfund/programs/clp/download/ism/ism1nfg.pdf.

EPA, 2015a. Hazardous and Solid Waste Management Systems; Management of Coal Combustion Residuals From Electric Utility, CFR Parts 257 and 261 , Federal Register, Vol. 80, No. 74, April 17, 2015.

EPA, 2015b. ProUCL Version 5.1.002 Technical Guide: Statistical Software for Environmental Applications for Data Sets with and without Nondetect Observations. EPA/600/R-07/041. October 2015.