TENNESSEE VALLEY AUTHORITY

Tennessee Valley Authority (TVA)

Paradise Fossil Plant (PAF) – Gypsum Disposal Area

40 CFR § 257.93(f)(6) Statistical Method Certification

REVISION LOG

Revision	Description	Date
0	Issued for Operating Record Posting	October 16, 2017
1	Statistical Method Certification revised to clarify approach to statistical analysis of groundwater monitoring data per the USEPA Unified Guidance for Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities (EPA 530/R-09-007, March 2009).	December 16, 2022

December 16, 2022

Tennessee Valley Authority
Paradise Fossil Plant
1101 Market Street
Chattanooga, Tennessee 37402

STATISTICAL METHODS CERTIFICATION 40 CFR § 257.93
TENNESSEE VALLEY AUTHORITY PARADISE FOSSIL PLANT
CCR GROUNDWATER MONITORING NETWORK
GYPSUM DISPOSAL AREA
UPDATE FROM 2017 CERTIFICATION

EPA'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.

This *Statistical Methods Certification* is provided for the Gypsum Disposal Area at the Tennessee Valley Authority (TVA) Paradise Fossil Plant (PAF), located in Drakesboro, Kentucky.

1.0 INTRODUCTION

This report includes a summary of the statistical methodology selected for evaluating groundwater monitoring data at PAF Gypsum Disposal Area and supports compliance with requirements outlined in 40 CFR § 257.93(f) and § 257.93(g) of the CCR Rule. To develop the most appropriate methods to validate assumptions, evaluate groundwater data, and develop background concentrations, the statistical methodology is based on *Statistical Analysis of Groundwater Data at RCRA Facilities, Unified Guidance*, March 2009, EPA 530/R-09-007 (Unified Guidance, 2009). This Statistical Methods Certification updates the prior version dated October 16, 2017.

Groundwater monitoring activities commenced in November 2016 (for initial baseline monitoring) and, at the time of this report, TVA obtained more than the minimally prescribed number of samples (i.e., "eight independent samples for each background and downgradient well") to comply with the initial baseline requirements included in § 257.90(b) of the CCR Rule.

This *Statistical Methods Certification* describes statistical methods applicable to detection monitoring, assessment monitoring, and corrective action. The statistical method for evaluating groundwater data in detection monitoring described in Section 3 of this document, prediction interval limits, is consistent with method/paragraph (3) of § 257.93(f), which includes a prediction interval procedure. In assessment monitoring or corrective action, the method described in Section 4 of this document, confidence intervals (and its variant, confidence bands), is consistent with *Unified Guidance* recommendations and is also justified under method/paragraph (5) of § 257.93(f), namely "Another statistical method that meets the performance standards of paragraph (g) of this section."

2.0 DEVELOPMENT OF BACKGROUND

2.1 Interwell Testing

When data from multiple upgradient, background wells are available, a determination will be made as to whether the upgradient data appear to come from the same population or whether there is evidence of statistically significant spatial variation at the facility. Data for each constituent will be plotted using box plots to assist in this determination, allowing concentrations within and across wells to be visualized. Analysis of Variance (ANOVA) will be utilized to evaluate whether spatial variation is statistically significant.

Conventionally, interwell statistical tests are used to evaluate whether compliance wells are consistent with, and in the expected range of, background. These tests are generally appropriate when there is no significant spatial variation at the site, and the natural groundwater gradient flows from the upgradient, background wells to the compliance locations.

The overarching goals of interwell testing will be to:

- Ensure that statistical comparisons will be adequately sensitive to detecting a facility release;
- Ensure that data used in testing reflect current background conditions; and
- Avoid confusing an impact caused by a release from the facility with a difference between wells caused by heterogeneous subsurface conditions.

2.2 Background Screening

Credible and adequate background data is the most important aspect to developing accurate and sensitive statistical limits. Standard parametric prediction and control chart limits for groundwater assume that the background data (1) are representative of current background conditions; (2) are statistically stable over time (i.e., not trending); (3) do not include (extreme) outliers; (4) include a sufficient number of samples to accurately estimate the variability in the underlying groundwater population, and thus be sensitive to a persistent change in groundwater concentrations; and (5) can be normalized, possibly via transformation. Non-parametric prediction limits, including rank-based and bootstrap methods, also rely on assumptions 1-4 (above), but do not require that the data be normalized (assumption 5, above).

To test these assumptions, any proposed background data will be screened prior to constructing statistical limits. Time series plots and formal trend tests will be used to check the stability of the data. The statistical pattern of the data along with the history and hydrogeology of the site will be used to gauge how well the data mimic current background conditions.

To handle potential outliers, one of two basic approaches will be utilized: (1) the standard method involves box plots and formal parametric outlier tests to identify, check for, and exclude any confirmed outliers, while (2) the robust method involves down-weighting of any potential outliers and the use of weighted, robust versions of standard statistical estimates (e.g., robust prediction limits) to curtail the influence of outlying values even when not formally excluded from the analysis. Robust methods have the advantage of bypassing sometimes uncertain judgments about whether specific observations are indeed outliers and can be adapted to cases where formal outlier testing is difficult, for instance, when the detection rate is low.

If average background concentration levels are changing over time (i.e., trending), the prospective background data may need to be truncated, removing older data to ensure that the resulting limits continue to represent current natural conditions. Confirmed outliers will either be flagged and de-selected from prospective background data prior to establishing statistical limits or will be down-weighted using alternate techniques robust to the presence of possible outliers, as discussed above. Any values flagged as confirmed outliers will be summarized in periodic reporting.

Probability plots and normality tests, adjusted for the presence of non-detects (Cameron, 2017), if any, will be used to identify and test best-fitting distributional models for the background data. If the data can be closely fit to a normal distribution (i.e., ‘normalized’), possibly via mathematical transformation, then a quasi-parametric bootstrap-t (DiCiccio & Efron, 1996) prediction limit or control chart will be constructed. The bootstrap-t technique uses bootstrapping on the normalized data, along with down-weighting of possible outliers, to more accurately estimate the kappa or h multiplier used to construct parametric prediction limits and control charts.

If the data cannot be normalized, a nonparametric rank-based or bootstrap prediction limit will be constructed instead. Non-parametric methods will also be considered when the skewness and pattern of the background data result in unrealistic and likely inaccurate parametric estimates.

The size of the background dataset impacts both the accuracy (false positive rate) and sensitivity (statistical power) associated with a prediction limit or control chart comparison. The CCR rule requires at least 8 baseline samples prior to the start of statistical analysis and evaluations, but often more background data is needed to meet EPA performance requirements for groundwater tests, especially at larger well networks. These requirements are discussed below (Section 3.1).

2.3 Periodic Updating of Background

Background data will be updated for interwell statistical limits by consolidating more recent sampling observations with historical background data at least every two years. Any new outliers in the combined background data will be either (1) flagged and removed, or (2) down-weighted prior to construction of statistical limits. This updating process will not only increase the background sample size but will also reduce the incidence of false positives when using nonparametric prediction limits and increase the statistical power of parametric prediction or control chart limits.

3.0 DETECTION MONITORING TESTS

Prediction limits and control charts are recommended by USEPA as primary techniques for detection monitoring. The detection monitoring methods described herein are in accordance with 40 CFR § 257.93(f)(3). Prediction limits and control charts both involve estimating statistical thresholds from background. If any new compliance observation exceeds the upper prediction limit or control chart limit, a potential statistical exceedance will be flagged. Retesting will then be conducted by collecting one or more independent resamples of the same well-constituent pair to confirm or disconfirm the initial exceedance. Any confirmed exceedance will be recorded as a statistically significant increase (SSI).

To conduct retesting, the pass one-of-m method, as described in the *Unified Guidance* (Chapter 19), allows for an efficient plan to confirm or disconfirm a potential SSI over background identified during detection monitoring. Depending on the background sample size, the target site-wide false positive rate, and the available time period in which to collect independent resamples, either a 1-of-2 or 1-of-3 method will be used when retesting is needed.

Under the CCR Rule, prediction limit and/or control chart tests will initially be implemented for all detected Appendix III parameters. Note that one parameter, pH, will require both upper and lower prediction and/or control chart limits. In that case, a potential SSI will be flagged whenever a new compliance measurement is either less than the lower statistical limit or higher than the upper statistical limit.

Parameters with all non-detects in background do not require formal testing but will be evaluated using USEPA's Double Quantification Rule (DQR). The DQR assumes that a significant change in groundwater quality has occurred whenever two consecutive detections of a parameter are observed after no previous detections. It is similar in nature to a nonparametric prediction limit with a single retest (1-of-2).

3.1 Statistical Performance Requirements

The *Unified Guidance* recommends two general criteria when designing a statistical detection monitoring program in order to meet statistical performance requirements under the Resource, Conservation and Recovery Act (RCRA) (and, by reference, the CCR Rule): (1) an annual site-wide false positive rate (SWFPR) of no more than 10%, and (2) statistical power of a site's 'weakest' test greater than or equal to the minimum benchmark power represented by the EPA reference power curves.

The first criterion informs the accuracy of statistical testing, limiting the occurrence of spurious (false) SSIs. The second criterion guides the sensitivity of testing, ensuring an adequate chance of identifying real changes in groundwater quality. In practical terms, the annual SWFPR is distributed evenly among the total number of well-constituent pairs and among the total number of statistical evaluations per year. Statistical limits will be constructed with sufficient background size and retesting in order not to exceed the per-pair portion of the overall false positive risk. Similarly, site-specific power curves associated with each distinct type of test will be constructed and compared to the EPA reference power curves to ensure adequate statistical power.

The CCR Rule indicates that if an SSI over background is confirmed for one or more Appendix III constituents during detection monitoring (that is, after all necessary retesting has been conducted), then the owner or operator of the CCR unit must, within 90 days: 1) establish an assessment monitoring program, 2) demonstrate that a source other than the CCR unit caused the SSI over background, or 3) demonstrate that the SSI over background resulted from error in sampling, analysis, statistical evaluation, or natural variation in groundwater quality. Written documentation must also be completed and certified by a qualified professional engineer within the 90-day timeframe.

4.0 ASSESSMENT MONITORING AND CORRECTIVE ACTION

The methods described herein for assessment monitoring or corrective action, confidence intervals (and its variant confidence bands), are consistent with *Unified Guidance* recommendations and are also justified under method/paragraph (5) of § 257.93(f), namely "Another statistical method that meets the performance standards of paragraph (g) of this section."

To implement assessment monitoring, the CCR Rule requires that all Appendix IV constituents be sampled, with any detected parameters added to the list of parameters sampled at least semiannually. To statistically evaluate these parameters for the CCR Unit, concentration data will be compared to Groundwater Protection Standards (GWPS) through the use of confidence intervals or their variant, confidence bands. A confidence interval is recommended and appropriate when the monitoring data do not exhibit a statistically significant trend. A confidence band is more appropriate when a trend is present. The GWPS for each constituent will be established as either the Maximum Contaminant Level (MCL), Regional Screening Level (RSL), where MCLs are not available, or as a statistical limit based on background when background concentrations are higher in concentration than the established MCL/RSL. On an annual basis, all Appendix IV parameters must be sampled, and newly detected parameters added to the list of parameters sampled at least semiannually.

4.1 Confidence Intervals

For each well-constituent pair, a trend test will be run to determine whether there is evidence of a significant trend. If not, a parametric confidence interval around the population mean will be constructed at the (one-sided) 99% confidence level when the compliance data follow a normal distribution.

If using a confidence interval approach, non-parametric bootstrap confidence intervals may be constructed if the data do not pass a normality test, due to skewness or other reasons. The accuracy of non-parametric intervals, including the bootstrap, depends in part on the number of observations used to construct the interval. When a well-constituent pair does not have sufficient sample size to ensure high statistical accuracy, a confidence interval with potentially less accuracy will be constructed but updated after each new sampling event until the desired accuracy is reached. The pair will also continue to be reported and tracked using time series plots and/or trend tests until enough data are available.

In assessment monitoring, a well is determined to be out of compliance, and has a statistically significant level (SSL), when the lower confidence limit (LCL), and thus the entire interval, exceeds the GWPS, as discussed in *Unified Guidance*. In the event of a statistical exceedance, assessment of corrective measures may be initiated within the time frame prescribed by the CCR Rule. Such remediation efforts will be evaluated via the continuing use of confidence intervals and/or confidence bands (Section 4.2) to determine remedial effectiveness.

4.2 Confidence Bands

If the compliance data at a given well-constituent pair show evidence of a significant trend, a linear regression line will be fit to the data and a confidence band with (one-sided) 99% confidence will be constructed around the trend line. Confidence bands will only be constructed on pairs with at least four independent samples.

To evaluate compliance with regulatory standards, the lower edge of the confidence band at the most recent sampling event will be compared to the GWPS. If the lower edge exceeds the GWPS at that point in time (thus guaranteeing the entire vertical cross-section of the band also exceeds the GWPS at that point), an SSL will be recorded. If the lower edge of the band does not exceed the GWPS, no SSL will have occurred. As new sampling events are collected, the trend estimate will be updated along with the confidence band.

4.3 Corrective Action

Should assessment of corrective measures be initiated, this information will be placed in the operating record and, if possible, an alternate source demonstration (ASD) will be made. If an SSL above GWPS is identified and an

ASD is not documented regarding any SSL above GWPS, efforts will be made to characterize the nature and extent of the release.

Once remediation activities begin, regular sampling will continue and confidence intervals and/or confidence bands will monitor the progress of remediation efforts. Confidence intervals and bands are compared to GWPS or other risk-based criteria to determine when clean-up levels are achieved.

Although in corrective action the same statistical techniques are used, the manner of the comparison is different from that in assessment monitoring. In corrective action a well-constituent pair is declared ‘clean’ when the entire confidence interval or cross-section of the confidence band at the most recent sampling event falls below a specified clean-up limit or GWPS (i.e., the upper confidence limit [UCL] or upper confidence band [UCB] falls below the clean-up or regulatory limit). This ‘clean’ status must be achieved for each of the Appendix IV parameters for a period of three consecutive years.

5.0 REFERENCES

Cameron, K (2017) ‘On-the-fly’ goodness of fit and outlier testing for left-censored data. In JSM Proceedings, Section on Statistics and the Environment, Alexandria, VA, American Statistical Association, 3445-53.

DiCiccio, TJ and Efron, B (1996) Bootstrap confidence intervals. Statistical Science, Vol 11, No 3, 189-228.

HDR (2017) Statistical Methods Certification for Compliance with the Final Coal Combustion Residuals Rule (40 CFR §257.93). October 16, 2017, 6 pp.

U.S. Environmental Protection Agency (2009) Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities: Unified Guidance. USEPA Office of Solid Waste, EPA 530-R-09-007.

Certification

This *Statistical Methods Certification*, Tennessee Valley Authority Paradise Fossil Plant Gypsum Disposal Area has been prepared in compliance with the United States Environmental Protection Agency coal combustion residual rule [40 Code of Federal Regulations (CFR) 257 Subpart D] by a qualified groundwater scientist or engineer with Golder Associates USA, Inc.

Golder Associates USA Inc.

Dawn L. Prell, CPG
Senior Hydrogeologist

Jeffery Frazier
Technical Principal, Geologist

I, Hunter Sartain, being a Professional Engineer in good standing in the State of Kentucky, do hereby certify to the best of my knowledge, information, and belief: (1) that the information contained in this certification is prepared in accordance with the accepted practice of engineering; (2) that the information contained herein is accurate as of the date of my signature; and (3) that in accordance with the requirements of 40 CFR §257.93, the selected statistical methods as described above in this certification are appropriate for evaluating the groundwater monitoring data for the Tennessee Valley Authority Paradise Fossil Plant's Gypsum Disposal Area located at 13246 State Route 176, Drakesboro, Kentucky, and that these methods meet the requirements described in 40 CFR § 257.93.

STATE OF KENTUCKY
HUNTER S. SARTAIN
37306
LICENSED
PROFESSIONAL ENGINEER

Hunter Sartain, PE
Kentucky Registered Professional Engineer No. 37306