COAL COMBUSTION RESIDUALS RULE
STATISTICAL METHODS CERTIFICATION
SOUTHERN ILLINOIS POWER COOPERATIVE (SIPC)

EMERY HOLDING POND
MARION POWER PLANT
MARION, ILLINOIS

Certification Statement 40 CFR § 257.93(f)(6) – Statistical Approach for the Evaluation of Groundwater Monitoring Data for the CCR Management Area

CCR Unit: SIPC: Marion Power Plant; Emery Holding Pond

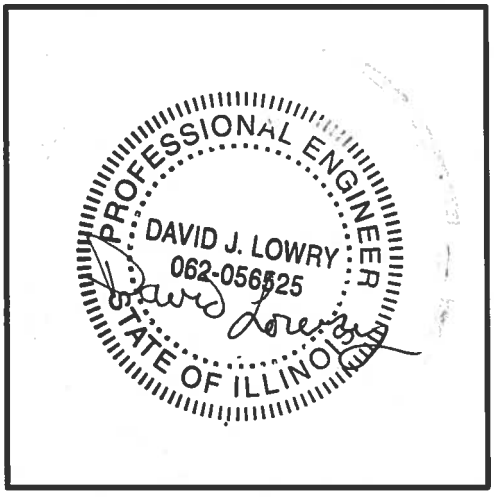
I, David Lowry, being a Registered Professional Engineer in good standing in the State of Illinois do hereby certify, to the best of my knowledge, information, and belief that the information contained in this certification is prepared in accordance with the accepted practice of engineering. I certify, for the above-referenced CCR Unit, that the statistical approach selected for the groundwater monitoring system, as described in this document, is appropriate for evaluating the groundwater monitoring data for the CCR management area. The statistical method(s) selected to evaluate the groundwater monitoring data for the CCR Unit is described below.

David Lowry

Printed Name

1/29/18

Date



Regulatory Guidance

Regulatory guidance provided in 40 CFR §257.90 specifies that a CCR groundwater monitoring program must include selection of the statistical procedures to be used for evaluating groundwater quality data as required by 40 CFR §257.93. Groundwater quality monitoring data has been collected under the detection monitoring program for the Emery Holding Pond (single unit CCR site) including analysis of eight independent groundwater samples from each background and downgradient well as required by 40 CFR §257.94(b).

40 CFR §257.93(f) outlines the statistical methods available to evaluate groundwater monitoring data. The statistical test(s) chosen will be conducted separately for each constituent in each monitoring well and will be appropriate for the constituent data and the data set distribution.

In accordance with 40 CFR §257.93(f)(6), a qualified professional engineer must certify that the selected statistical method is appropriate for evaluating the groundwater monitoring data for the CCR unit.

Statistical Analysis Approach

When conducting statistical evaluations of groundwater data sets, it is most prudent to use a suite of statistical methods that are dependent on the character of the data and their distributions. For the groundwater data collected from the groundwater monitoring system at the Emery Holding Pond, the statistical analyses will be based on an interwell approach for the purpose of determining if a CCR unit(s) has caused a statistically significant increase. The single unit groundwater monitoring system contains one upgradient and four downgradient wells that are installed in the uppermost aquifer; therefore, an interwell approach is considered appropriate. The statistical algorithms used for the interwell approach will be chosen based on the groundwater constituent data and their distributions as well as consideration of natural seasonally- or spatially-varying groundwater constituent concentrations.

Eight rounds of baseline groundwater monitoring data were collected and analyzed for the 40 CFR § 257 Appendices III and IV constituents.

A preliminary, exploratory statistical analysis was performed on the eight rounds of baseline data to initially assess the constituent data to determine the most appropriate statistical approach(es) for the data. The data was examined for outliers and the percentage of non-detect values to verify that the data collected are suitable for statistical analysis. The data was also examined using goodness-of-fit tests to determine the most appropriate statistical distribution, time series plots, and areal maps to determine if

seasonal or spatial variations in constituent concentrations are present. Based on this preliminary evaluation of the data, the final statistical approach selected is deemed appropriate for evaluating groundwater in accordance with the CCR rule.

The statistical approach for each detection monitoring event is summarized in Figure 1 below. All potentially applicable statistical methods are described in the following paragraphs in the event that any changes in data distributions or non-detect percentages occur as the dataset increases with future sampling events

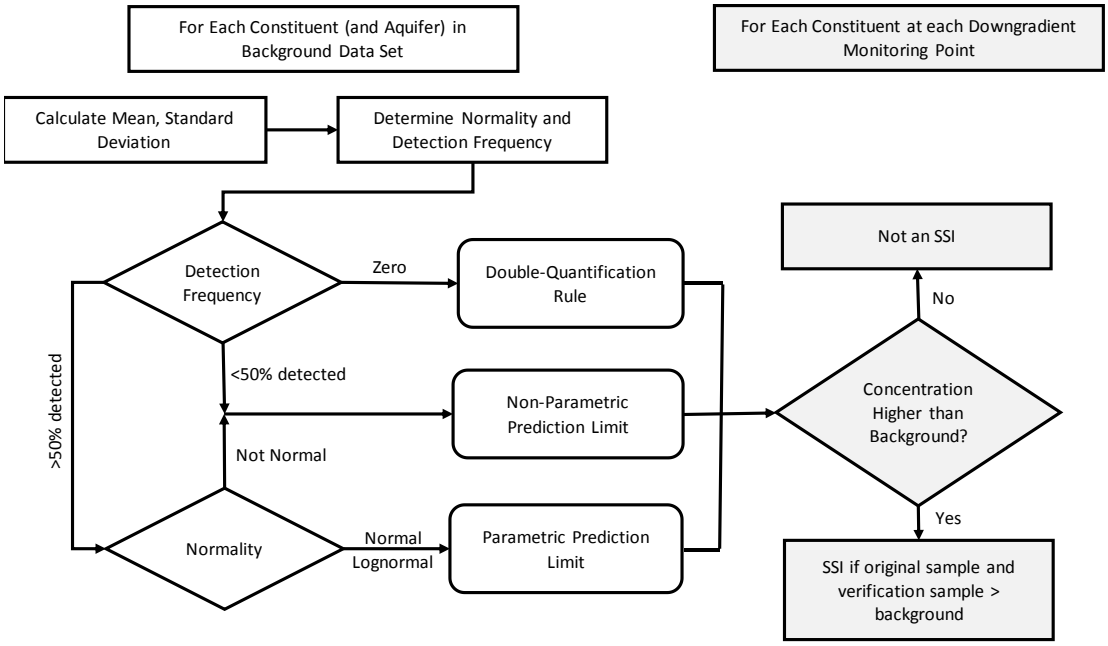


Figure 1. Flow chart for Detection Monitoring statistical evaluation (from EPRI, 2015).

Shapiro-Wilk W Test for Normality and Lognormality

The type of data distribution is required to be determined in order to select an appropriate statistical method [per CCR Rule 40 CFR 257.93(g)(1)]. The Shapiro-Wilk W test is a goodness-of-fit test (two-sided and parametric) on whether the data have been drawn from an underlying normal distribution (Gilbert, 1987). The null hypothesis H_o is that the population has a normal distribution. The alternative hypothesis H_a is that the population does not have a normal distribution. A goodness-of-fit test for lognormal distributions is performed by first taking the natural logarithm of all the data values and then applying

the W test on the transformed data. The Shapiro-Wilk W test is valid for sample sizes less than or equal to 50 values. For data sets larger than 50, the Shapiro-Francia test is used (Gibbons, 1994).

Parametric Prediction Intervals for Future Compliance Values

The prediction interval method is one of the statistical methods cited in the CCR Rule [40 CFR 257.93(f)(3)]. Both parametric and non-parametric versions of this statistical test are available (as explained in the Unified Guidance USEPA, 2009), which is cited in the discussion section of the CCR Rule [p. 21401 K(3) and other places]. The parametric prediction interval method calculates upper and lower values, based on background data, against which future values from compliance locations will be compared (USEPA, 1989). This method calculates a parametric prediction interval from all pooled background data for a specified base period from one or more locations. The data are then used to compute a prediction interval for an initial period. The parameter value for each of the compliance location intervals is then compared to the upper bound of the prediction interval. A statistically significant exceedance time period is indicated when the value of an individual measurement for a compliance location exceeds the upper bound of the prediction interval, or the lower bound for pH.

The data or transformed data should be normally distributed. A minimum of four observations per period are recommended for the compliance location data. A minimum of a one year base period of background observations is recommended for construction of the prediction interval. The data should be free of outliers.

Non-Parametric Prediction Interval for Future Compliance Values

The prediction interval method is one of the statistical methods cited in the CCR Rule [40 CFR 257.93(f)(3)]. The non-parametric prediction interval calculates the prediction interval using pooled background data over a specified base period. The background data are pooled from one or more locations. The pooled background data are ranked and the minimum value is identified as the one-sided, lower prediction limit for pH only, P_l, and an appropriate value is identified as the one-sided, upper prediction limit, P_u, depending on the number of background samples (as described in Section 18.3.1 USEPA, 2009). Lower and upper, non-parametric, one-sided confidence limits are computed for the compliance locations. No assumption is made concerning the underlying distribution of the data. However, the assumption is made that the unknown distribution in the background and compliance data is continuous and is the same in both background and compliance datasets in the absence of contamination.

At least four background values and at least one compliance location are needed for this analysis. However, there need not be any actual data in the selected compliance locations if the user only wishes to determine the prediction intervals. If an individual measurement from a compliance location exceeds

the upper limit, then a statistically significant exceedance is declared. Normally, only an exceedance of the upper limit is of concern, except for pH. A general discussion of estimating non-parametric prediction limits and alternative verification procedures is given in Gibbons (1994) and in Section 18.3.1 (USEPA, 2009).

Non-Parametric Prediction Interval for Future Compliance Median

The non-parametric prediction interval method is one of the statistical methods cited in the CCR Rule [40 CFR 257.93(f)(3)]. The USEPA (2009) describes in Section 18.3.1 of the Unified Guidance the various strategies available for setting the upper prediction limit when the background data are non-parametrically distributed and sufficient compliance data are available. In particular, the option of using the median of three future compliance measurements to test against the upper prediction limit is described on page 18-21. For that approach, the user is given the option of setting the upper prediction limit to either the largest, the 2nd-to-largest, or the 3rd-to-largest background measurement. The corresponding confidence limit for each of these choices is affected by the background sample size n_{bg}. The confidence level increases as n_{bg} increases. In addition, for the same sample size n_{bg}, the confidence level decreases as one selects values smaller than the maximum when the prediction limit is chosen to be the jth largest background measurement. A complete statistical table is given on page D-31 of the Unified Guidance (USEPA, 2009). Note that for the 95% confidence level, only 9 background data values are needed when selecting the maximum background measurement as the upper prediction limit, as compared to needing 24 background values when selecting the third-to-largest background measurement for the upper prediction limit.

Non-parametric Poisson Prediction Interval

The non-parametric prediction interval method is one of the statistical methods cited in the CCR Rule [40 CFR 257.93(f)(3)]. The Poisson prediction interval method calculates upper and lower, one-or two-sided, non-parametric prediction limits, based on background data, against which future data from compliance wells will be compared. The Poisson distribution in statistics is used to model rare events. The Poisson model describes the behavior of a series of independent events that occur while taking a large number of observations. For the purposes of this document, an *event* occurs when the chemical concentration of a sample is above the level of detection. The probability of detection is low but it remains constant from observation to observation.

One of the key distinctions between the Poisson model and other non-parametric models is that the Poisson model utilizes the magnitude of the measured concentrations in its algorithm. Upon selecting a *scaling* parameter, all sampled concentrations for a particular chemical at a location are then converted into an equivalent number of *chemical units* or *counts*. The model then computes the average rate of

occurrence of these counts for the chemical from a specified sample set. Finally, it predicts the lower and upper limit for an interval that will contain all of the future measurements of this chemical at the location.

The Poisson model can only be used if there is available at least one background measurement that is detected. As discussed in the Unified Guidance (USEPA, 2009, pp. 6-11 and 6-37), the Double Quantification Rule must be used when none of the background measurements are detected.

Double Quantification Rule

The Double Quantification Rule applies when all data from the background wells have no detected values for a particular constituent. If, during a sampling event, that particular constituent is detected in a downgradient well, a subsequent sample (resampling) would be collected from that well and analyzed. If the downgradient concentration for that constituent in that given well is higher than the reporting limit in both the original sample *and* in the verification resample, then a statistically significant increase determination would be made.

Statistical Methods for Non-Detect Values Less than 15 Percent

Additional statistical analysis methods may be applicable to upgradient and downgradient wells when non-detect values are less than 15 percent as described below.

Behrens-Fisher Student t-Analysis

The Student t-test is a one-sided, parametric test that compares the means from two data sets. If confidence ranges for the means overlap, then the two means are not significantly different. This test assumes normally-distributed data.

Satterthwaite's t-test (Iman and Conover, 1983) is a modified form of the standard t-test that is appropriate when the background and site distributions have unequal variances. Testing data sets with unequal variances are called Behrens-Fisher problems. The Student t-test makes three key assumptions: (1) that the two location data sets are independent; (2) not serially correlated; and (3) that both location data sets have normal distributions (Guttman et al.,1971; Gilbert, 1987). If these assumptions are not met, the Wilcoxon Rank-Sum test should be used for determining whether the means of two locations are different (Loftis, et al., 1987).

Wilcoxon Rank-Sum Analysis

The Wilcoxon Rank-Sum test is a one-sided, non-parametric test that compares the means from two data series. This method is an alternative statistical test method allowed under the CCR Rule [40 CFR

257.93(f)(5)]. If confidence ranges for the means overlap, then the two means are not significantly different. If multiple background locations are specified, they are pooled. The evaluation is conducted for each compliance location/parameter combination, and determines whether the mean concentration of the specified parameter at the compliance location is statistically higher than the mean concentration of that parameter at the pooled background locations. The Wilcoxon Rank-Sum test assumes that: (1) both data sets contain random values from their respective populations, and (2) in addition to independence within each data set, there is mutual independence between the two sample sets. No assumptions are made about data distribution. The null hypothesis is that the two location means are equal, and the alternative hypothesis is that the two location means are different.