STATISTICALLY SIGNIFICANT EXCEEDANCE- UNDERSTANDING FALSE POSITIVE ERROR

Arun Kammari[1]

[1]Haley & Aldrich, 70 Blanchard Road, Suite 204, Burlington, MA 01803-5100

ABSTRACT

Under the newly enacted Coal Combustion Residual (CCR) Rule, CCR impoundments and lateral expansions of CCR units are subject to detection monitoring requirements to evaluate the groundwater data for determining Statistically Significant Increase (SSI) over background levels for each of the Appendix III constituents. The presence of SSI is evaluated using one of the five statistical methods complying with the test performance standards prescribed in the CCR rule. When evaluating the statistical methods by these performance standards, it is important to understand the ability of a chosen statistical method as it relates to minimizing the occurrence of false positives (i.e., minimizing the risk of falsely declaring a site to be out-of-compliance). A systematic application of basic statistical principles is crucial for balancing false positive errors in designing good testing procedures.

INTRODUCTION

As mandated by the newly enforced CCR rule, the existing or new CCR landfills and CCR impoundments are required to monitor and evaluate (i.e. detection monitoring program) the Appendix III constituents by establishing an effective monitoring groundwater well network surrounding a CCR management unit to evaluate the potential impacts of CCR units and statistically determine if the impact requires assessment and corrective actions by the regulated facility owner or operator. Groundwater detection monitoring involves statistical comparison between site specific background groundwater data and downgradient compliance groundwater data to identify a real release from CCR units to groundwater when it occurs. The success of a detection monitoring program is dependent on the following factors:

1. Installation of an effective monitoring well network and establishing appropriate background data pool
2. Choosing an appropriate statistical method and testing of underlying statistical assumptions and statistical design considerations

APPROPRIATE BACKGROUND DATA POOL

Establishment of appropriate background data depends on factors such as choosing an effective monitoring well network and selecting a statistical approach (e.g., inter-well vs. intra-well). The installation of an effective monitoring well network is a key and important first step in correctly identifying the target background population and minimizing the false positives (falsely declaring a site to be out of compliance) as many of these constituents are naturally occurring in the soil/groundwater and vary substantially across the site due to natural geochemical factors. The selection of an appropriate monitoring network (§257.91) should be based on development of the conceptual site model (CSM) and refinement of the model when new information is gathered. The CCR rule specifies that at least one upgradient and three downgradient wells in the uppermost aquifer at the facility is minimum to perform the groundwater monitoring at the facility assuming a definable groundwater gradient exists as shown in Figure 1.

Figure 1: Typical CCR Unit Monitoring Network Example
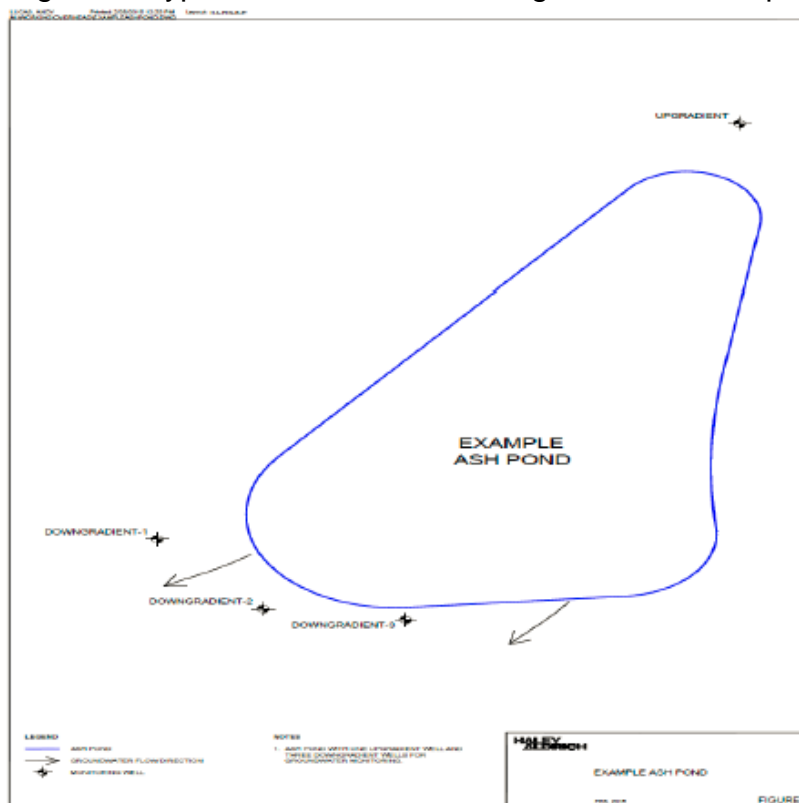
In considerable cases, establishment of a definable groundwater gradient is problematic due to many factors, such as slow groundwater flow, seasonal change in groundwater gradient, and groundwater mounding. Selection of appropriate background differs between the following two statistical approaches:

Approach-1: An inter-well approach (comparison between upgradient and downgradient monitoring network) is recommended:

- If up-gradient and down-gradient well measurements are comparable and have common variance (i.e. drawn from the same statistical population).
- If the groundwater that flows beneath the regulated unit follow a definable pathway from upgradient to downgradient wells
- If the underlying aquifer is continuous
- If the up-gradient and downgradient wells are screened at the same hydrostratigraphy

Approach-2: An intra-well approach (comparison between past and present data within a given compliance well) is recommended:

- If the site hydrogeology suggests heterogeneity or spatial variability
- If more than one aquifer underlies the regulated unit or if the up-gradient and downgradient wells are screened at different aquifers
- If the groundwater flow path way cannot be determined due to complex hydrogeological settings

In the example shown in Figure 2, the onsite wells show a significant spatial variation using Levene's test and box plot comparison. In this case, choosing an inter-well approach might lead to false positive conclusions. Therefore, an intra-well approach might be an appropriate statistical approach to establish the background data pool.

Figure 2: Equality of Variance and Box Plot Comparison
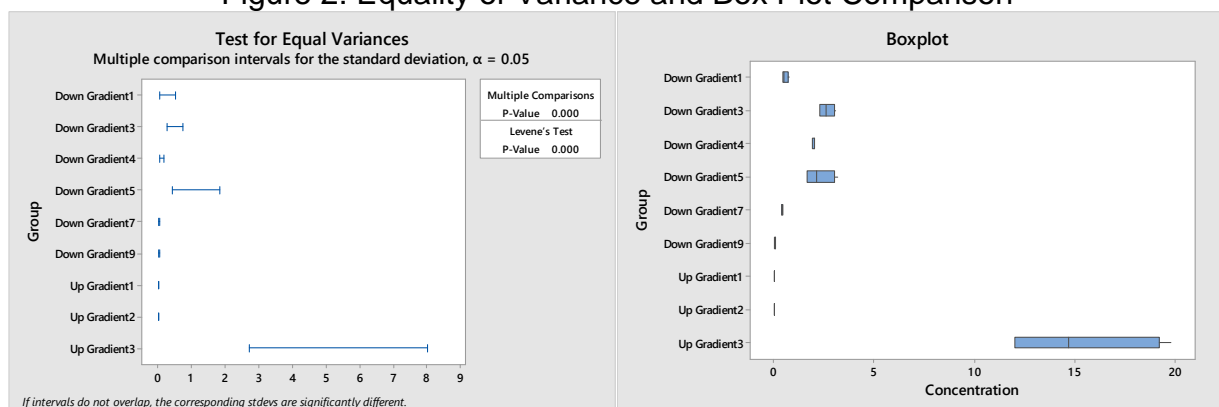
Choosing a correct statistical comparison approach is an initial investment that can save significant money and time and prevent poor decisions.

APPROPRIATE STATISTICAL METHOD

Once the background data and statistical approach are identified, the next important step in the detection monitoring phase is the selection of appropriate statistical methods to determine if there is a statistically significant increase (SSI) in the constituent at the downgradient locations. The selection of statistical methods addressing groundwater detection monitoring (40 CFR § 257.93 and 257.94) in the CCR rule follows the methods prescribed in the "Statistical Analysis of Groundwater monitoring data at RCRA facilities unified guidance". The CCR Rule prescribed five different statistical methods and the tests to comply with the performance standards (§257.93(g)(1-6)) as presented below for detection monitoring evaluation.

1. A parametric analysis of variance (ANOVA) followed by multiple comparison procedures
2. An analysis of variance based on ranks (non-parametric ANOVA) followed by multiple comparison procedures
3. A tolerance interval method or prediction interval method;
4. A control chart method
5. Another statistical method that meets the performance standards prescribed in §257.93(g)(1-6)

Selection of an appropriate statistical method is dependent on underlying statistical assumptions and statistical design considerations. Testing of statistical assumptions are a crucial and mandatory step in the selection of an appropriate testing method and minimizing false positives. Most importantly, all relevant statistical assumptions must be evaluated using Exploratory Data Analysis (EDA) methods. EDA testing methods includes descriptive numerical summary statistics such as measures of centrality (mean, median), measures of spread (standard deviation, variance, interquartile range), and measures of shape (skewness and kurtosis), as well as graphical plots such as histograms, box plots, scatter plots, time series plots, and probability plots. EDA methods helps to identify whether the underlying assumptions of statistical methods are met. A thorough testing and confirmation of statistical assumptions are essential in minimizing the risk of falsely declaring a site to be out-of-compliance. Common statistical assumptions are:

- Normality test
- Outliers test
- Background stability evaluation
- Spatial variation evaluation
- Temporal independence
- Nondetects imputing

Figure 3 presents a few graphical plots of an example dataset and Table 1 presents some of the descriptive statistics of the data pool.

Figure 3: Histogram, Dotplot and Tim-series Plots of Groundwater Data

Table -1 : Descriptive statistics of the Groundwater data pool

| Variable	Frequency of Detection			Percent Non-Detects	Range of Non-Detects	Mean	50th Percentile (Median)	95th Percentile	Maximum Detect	Variance	Standard Deviation	Coefficient of Variation
Appendix III: mg/L												
Downgradient-1	8	/	8	0%	N/A	0.547	0.508	0.717	0.75	2.04E-02	0.143	0.261
Downgradient-3	8	/	8	0%	N/A	2.635	2.6	3.038	3.08	1.40E-01	0.375	0.142
Downgradient-4	8	/	8	0%	N/A	1.983	1.995	2.034	2.04	3.49E-03	0.0591	0.0298
Downgradient-5	8	/	8	0%	N/A	2.25	2.11	3.081	3.18	5.54E-01	0.744	0.331
Downgradient-7	8	/	8	0%	N/A	0.406	0.405	0.433	0.436	6.67E-04	0.0258	0.0636
Downgradient-9	8	/	8	0%	N/A	0.0564	0.0551	0.0788	0.081	4.46E-04	0.0211	0.375
Upgradient-1	8	/	8	0%	N/A	0.0228	0.0231	0.0303	0.031	5.43E-05	0.00737	0.323
Upgradient-2	8	/	8	0%	N/A	0.0284	0.0307	0.0342	0.0342	5.81E-05	0.00762	0.269
Upgradient-3	8	/	8	0%	N/A	15.3	14.7	19.44	19.8	1.55+01	3.934	0.257

APPROPRIATE STATISTICAL DESIGN CONSIDERATIONS

Apart from the statistical assumptions, statistical design considerations are also crucial in minimizing false positive errors and choosing proper test procedures. Every decision, whether its statistically concluded or non-statistically concluded, involves certain amount of uncertainty in decision-making. Prior understanding of statistical design considerations provides clear understanding of the anticipated uncertainty and gives the ability to adjust the design when the anticipated uncertainty is unacceptable. The statistical design is measured based on the test confidence level (1-α), expected false positive (α) and false negative (β) rates or statistical power (1-β). A well-designed evaluation provides a higher confidence level on the chosen test procedure; and, controls the expected false positive and false negative rates or statistical power of the test. Figure 4 presents these factors in a hypothesis testing schema.

Figure 4: Hypothesis Testing Schema
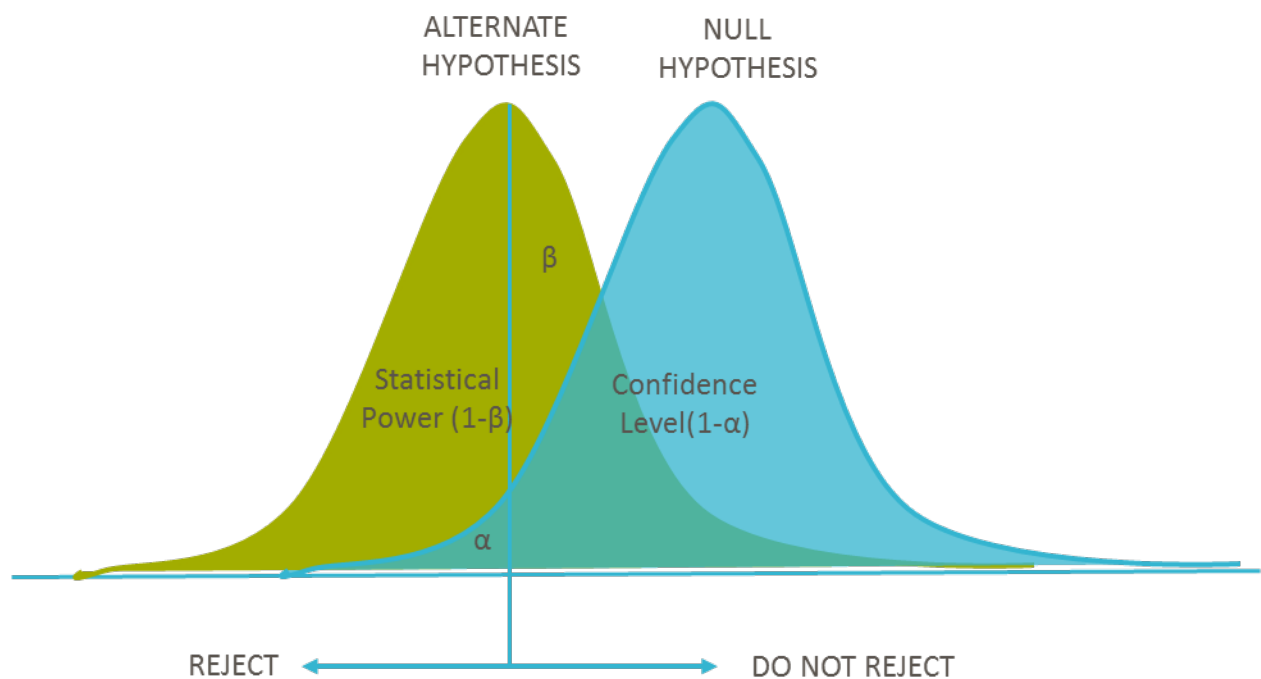
In statistical hypothesis testing, a false positive (also called Type-I error or Alpha) occurs when the null hypothesis about a population is rejected when it is actually true. A false negative (also called Type II error, or beta) refers to failing to reject the null hypothesis or conclusion when it is actually false. The flow chart in Figure-5 illustrates the Type-I and Type-II error during the test of hypothesis.

Figure 5: Hypothesis Test Errors
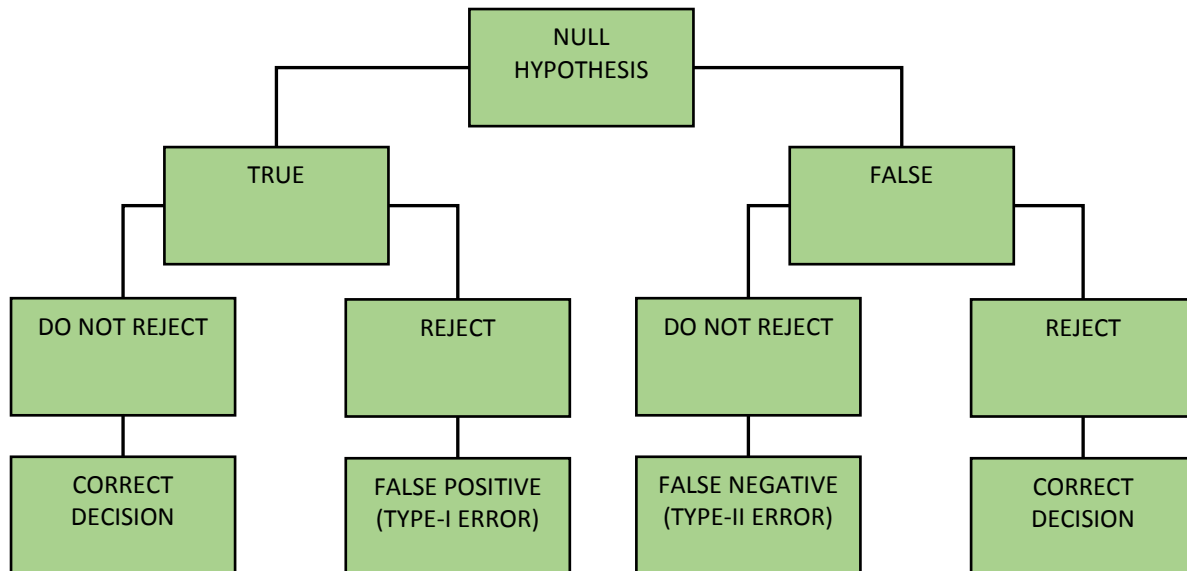
In CCR groundwater evaluation, the null hypothesis typically refers to a site groundwater that is not impacted (no SSI) by the CCR unit and the alternate hypothesis refers to site groundwater that is impacted by the CCR unit. A false positive error means if the statistical evaluation erroneously concludes that the groundwater is impacted when it is actually not impacted. The occurrence of false positives can be minimized by constructing the statistical test with a high confidence level. A confidence level (1-alpha) is the probability of not committing a false positive error, i.e. a high probability of correctly deciding that site is not impacted when it is actually true. The CCR rule prescribes that "*If an individual well comparison procedure is used to compare an individual compliance well constituent concentration with background constituent concentrations or a groundwater protection standard, the test shall be done at a Type I error level no less than 0.01 for each testing period. If a multiple comparison procedure is used, the Type I experiment wise error rate for each testing period shall be no less than 0.05; however, the Type I error of no less than 0.01 for individual well comparisons must be maintained.".* Since CCR groundwater monitoring generally involves testing of multiple chemicals at multiple sampling points that involves multiple comparison procedure, the probability that at least one of those tests will falsely indicate a significant result is much higher than the individual test false positive rate. This alternate probability of error is known as the site-wide false positive rate (SWFPR). The Unified Guidance, recommends designing the detection monitoring program with an annual cumulative SWFPR of 10% to control the number of false positive decision to minimum, regardless of the number of individual statistical tests that are run each year. The SWFPR is further explained in the example below.

In the example, seven Appendix-III constituents at six downgradient well locations statistically evaluated semi- annually constitute 84 annual tests (7x6x2). To maintain

10% cumulative annual SWFPR, the pre-test false positive should be maintained at "0.0013" as computed below.

$$SWFPR = 1 - (1 - alpha)^{\text{no. of tests}}$$
$$alpha = 1 - (1 - SWFPR)^{1/(\text{no. of tests})}$$
$$alpha = 1 - (1 - 0.1)^{1/(84)}$$
$$alpha = 1 - (0.9)^{0.012}$$
$$alpha = 0.0013$$

If you consider alpha of 0.01 for each individual test then the cumulative annual SWFPR is going to be 57% for the site. Often, the significance level of a test is equated with the false positive rate.

REFERENCES

[1] EPA (United States Environmental Protection Agency). 2009. Statistical Analysis of GroundWater Monitoring Data at RCRA Facilities, Unified Guidance. Washington, DC: EPA. EPA 530/R-09-007.

[2] Iswar P. Murarka. Detection Monitoring under the USEPA CCR Regulations and Statistical Evaluation of the Resulting Data, World of Coal Ash, 2015

[3] ITRC (Interstate Technology & Regulatory Council). 2013. Groundwater Statistics and Monitoring Compliance, Statistical Tools for the Project Life Cycle. GSMC-1. Washington, D.C.: Interstate Technology & Regulatory Council, Groundwater Statistics and Monitoring Compliance Team. http://www.itrcweb.org/gsmc-1/.

[4] The Digital Textbook. www.statlect.com