GUIDELINE 2 - STATISTICAL ANALYSIS OF GROUNDWATER MONITORING DATA FROM SOLID WASTE MANAGEMENT FACILITIES

North Dakota Department of Environmental Quality - Division of Waste Management
4201 Normandy St., Bismarck, ND 58503-1324
Telephone: 701-328-5166 ● Fax: 701-328-5200 ● Email: solidwaste@nd.gov
Website: https://deq.nd.gov/wm
Updated 03-2023

I. Introduction

This document has been prepared by the North Dakota Department of Environmental Quality (Department) for the purpose of assisting owners and operators who conduct groundwater monitoring to fulfill regulatory and permit requirements at solid waste facilities. It is not intended to provide detailed instruction on environmental statistics. It is recommended that the evaluation of groundwater monitoring results be performed by qualified environmental professionals who have experience in groundwater monitoring statistics. Questions and comments are welcome and can be addressed to the Department's Division of Waste Management.

The owner or operator of a solid waste management facility must indicate the statistical method(s) that will be used in the analysis of groundwater monitoring results in a groundwater monitoring plan (North Dakota Administrative Code (NDAC) Sections 33.1-20-08-06 and 33.1-20-13-02). The number of samples collected, and the frequency of collection, must be consistent with the statistical method selected.

Statistical analysis methods are applicable to all existing units, new units, and lateral expansions of existing units that are required to conduct groundwater monitoring. The use of statistical methods to evaluate monitoring data is necessary for the duration of the monitoring program including the post-closure period.

Several options for analysis of groundwater data are provided. Other methods are not excluded if they can be shown to meet statistical performance standards. The recommended methods include both parametric and non-parametric procedures which differ primarily in constraints placed by the statistical distribution of the data. Control chart, tolerance interval, and prediction interval approaches may also be applied.

The owner or operator must conduct the statistical comparisons between upgradient and downgradient wells within a reasonable period of time after completion of each sampling event and receipt of validated data (NDAC Section 33.1-20-13-02). For coal combustion residuals (CCR) facilities, the statistical comparisons must be performed within ninety days. The statistical procedure must conform to the performance standards discussed in Section II. Procedures to treat data that are below analytical method detection levels and seasonality effects are necessary prior to statistical analysis.

II. Technical Considerations

<u>Methods</u>

In most cases, the Department recommends one of five statistical methods for evaluating groundwater monitoring data. For CCR facilities these methods are listed in NDAC Subdivision 33.1-20-08-06(3)(f). Different methods may be selected for each groundwater quality constituent.

The appropriateness of a method must be substantiated by demonstrating that the distribution of the data for that constituent is appropriate for the method. Selection and use of a specific method is described in the U.S. Environmental Protection Agency's (USEPA) "Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities - Unified Guidance" (USEPA, 2009) (hereinafter referred to as "Unified Guidance".) The methods include the following:

1. A parametric analysis of variance followed by multiple comparison procedures to identify statistically significant evidence of contamination. The method must include estimation and testing of the contrasts between each compliance well's mean and the background mean levels for each constituent.

2. An analysis of variance based on ranks followed by multiple comparison procedures to identify statistically significant evidence of contamination. The method must include estimation and testing of the contrasts between each compliance well's median and the background median levels for each constituent.

3. A tolerance or prediction interval procedure, in which an interval for each constituent is established from the distribution of the background data and the level of each constituent in each compliance well is compared to the upper tolerance or prediction limit.

4. A control chart approach that gives control limits for each constituent.

5. An alternative statistical method that meets the performance standards for the other methods.

These statistical analysis methods are necessary to determine whether a significant increase over background has occurred. The statistical analysis of monitoring data occurs after receiving validated results from each sampling and analysis event. If an alternative method is used, then the Department must be notified and a justification for its use must be provided.

Performance Standards

Performance standards are necessary for a statistical analysis method to provide reliable and valid conclusions. These standards are:

1. The statistical method used to evaluate ground water monitoring data must be appropriate for the distribution of constituents. Normal distributions of data values must use parametric methods. Nonnormal distributions must use nonparametric methods. If the distribution of the constituents is shown to be inappropriate for a normal theory test, then the data must be transformed, or a distribution-free (nonparametric) theory test must be used. If the distributions for the constituents differ, more than one statistical method may be needed.

2. If an individual well comparison procedure is used to compare an individual compliance well constituent concentration with background constituent concentrations or a ground water protection standard, the test must 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 must be no less than 0.05; however, the type I error of no less than 0.01 for individual well comparisons must be maintained. This performance standard does not apply to tolerance intervals, prediction intervals, or control charts.

3. If a control chart approach is used to evaluate ground water monitoring data, the specific type of control chart and its associated parameter values must be such that this approach is at

least as effective as any other approach in this section for evaluating ground water data. The parameter values must be determined after considering the number of samples in the background database, the data distribution, and the range of the concentration values for each constituent of concern.

4. If a tolerance interval or a predictional interval is used to evaluate ground water monitoring data, the levels of confidence and, for tolerance intervals, the percentage of the population that the interval must contain, must be such that this approach is at least as effective as any other approach in this section for evaluating ground water data. These parameters must be determined after considering the number of samples in the background database, the data distribution, and the range of the concentration values for each constituent of concern.

5. The statistical method must account for data below the limit of detection with one or more statistical procedures that shall be at least as effective as any other approach in this section for evaluating ground water data. Any practical quantization limit that is used in the statistical method must be the lowest concentration level that can be reliably achieved within specified limits of precision and accuracy during routine laboratory operating conditions that are available to the facility.

6. If necessary, the statistical method must include procedures to control or correct for seasonal and spatial variability as well as temporal correlation in the data.

The statistical performance standards provide a means to limit the possibility of making false conclusions from the monitoring data. The specified error level of 0.10 for individual well comparisons for probability of Type I error (indication of contamination when it is not present, or false positive) essentially means that the analysis is predicting with 90 percent confidence that no significant increase in contaminant levels is evident. The corollary is that there is only a 10 percent chance that a Type II error (failure to detect a significant increase in constituent concentration, or false negative) has occurred.

Nondetected results must be treated in an appropriate manner or their influence on the statistical method may invalidate the statistical conclusion. Further discussion of nondetected results is found later in this section. In addition to the statistical guidance provided in the USEPA Unified Guidance, the following references may be useful for selecting other methods: Dixon and Massey, 1969; Gibbons, 1976; Aitchison and Brown, 1957; and Gilbert, 1987.

III. Multiple Well Comparisons

The goal of monitoring well placement is to provide representative samples of water quality both upgradient and downgradient of the waste management unit. In this manner, changes in the quality of groundwater can be determined. If more than two wells (upgradient and downgradient combined) are screened in the same stratigraphic unit, then the appropriate comparison method is a multiple well comparison using analysis of variance or ANOVA. The analysis of variance test compares the average concentration among wells to determine if they are from the same continuous distribution. The ANOVA test includes both parametric and non-parametric procedures. Chemical data from groundwater tends to follow a log normal distribution and data usually needs to be transformed prior to applying a parametric ANOVA procedure.

By making a log transformation, data will generally be converted to a normal distribution. By applying a chi-squared procedure, probability plots or other normality tests on the residuals (errors) from the ANOVA procedure, the normality of the transformed data can be determined. In addition, the variance of data from each well in the comparison must be approximately equivalent; this condition can be checked using Bartlett's test. Both Bartlett's test and four normality tests are presented in the USEPA Unified Guidance.

If the transformed test data does not conform to the normality assumption, a non-parametric ANOVA procedure should be used. The non-parametric statistical procedures are not dependent on the mathematical properties of a specified distribution. The non-parametric equivalent to the parametric ANOVA is called the Kruskall-Wallis Test which uses ranking methods to compare the data.

If the data displays seasonality (consistent time dependent increases or decreases in parameter values), a two-way ANOVA procedure should be used. If the seasonality can be corrected, a one-way ANOVA procedure may still be appropriate. Methods to treat seasonality are described in the USEPA Unified Guidance.

ANOVA procedures determine whether different wells have significantly different concentrations of constituents. ANOVA procedures are followed by multiple comparisons procedures that are used to discriminate between wells showing significant differences. The multiple comparison procedures test the contrast between the mean value of each monitoring parameter for the background to the compliance boundary well. For limitations on multiple comparison tests when more than five wells are used, refer to the USEPA Unified Guidance.

If data lacks variability between upgradient and downgradient wells, other methods, including tolerance intervals and prediction intervals, may be used in place of analysis of variance. These methods may be especially appropriate where the site geology is very homogeneous. Both methods are similar and consist of constructing confidence intervals about the mean of the background well data. The confidence interval should be one-sided since the objective of monitoring is to detect significant changes that are greater than the background mean. The confidence level should be set to 90 percent. These methods are discussed also in the USEPA Unified Guidance.

IV. Individual Well Comparisons

Sites with complex geology may have stratigraphic units which are monitored by only one background (upgradient) and one downgradient well. Under these conditions the ANOVA procedure, which requires a minimum of three wells, may not be used. In these instances, other statistical methods would be required to compare background and compliance boundary data.

A number of statistical procedures can be found in statistic references, for the comparison of means where variances are equal or where variances are unequal. The distribution assumptions of the method selected must be satisfied by the distribution of the data, and the confidence level should be no less than 90 percent for a Type I error.

V. Intra-Well Comparisons

Intra-well comparisons, where data of one well is evaluated over time, are useful in evaluating trends in individual wells and for identifying seasonal effects in the data. During initial rounds of groundwater monitoring, seasonal effects should not be misconstrued for either positive or negative trends.

Some existing facilities may not have valid upgradient background data. Intra-well comparisons may represent the only valid comparison available. For example, a facility constructed on a hilltop with radial groundwater flow around the facility, may not have an upgradient well. If the geology is not similar nearby, it may not be able to clearly establish upgradient background groundwater quality from adjacent wells.

A significant positive trend in the data from a downgradient monitoring well may provide sufficient evidence to conclude that the landfill unit is affecting the well. Such trends should be evident in several parameters since contaminant migration from a landfill will result in an influx of multiple constituents into the groundwater flow system.

Control charts may be used for intra-well comparisons but are only appropriate for uncontaminated wells. If a well is intercepting a release, then it is already in an out-of-control state which violates the principal assumption underlying control chart procedures.

Time series analysis (plotting concentrations over time) is extremely useful for identifying trends in monitoring data. Such data may be adjusted for seasonality effects to aid in assessing the degree of change over time. Guidance for intra-well comparison techniques and limitations of the techniques are provided in the USEPA Unified Guidance.

VI. Treatment of Nondetections

The treatment of data below the detection limit of the laboratory's analytical method is dependent upon the number or percentage of nondetections (NDs) and the statistical method employed. If a large proportion (greater than 50 percent) are present, a professional statistician should be consulted. The USEPA Unified Guidance provides more information on this topic which is briefly summarized here.

If the amount of data below detection is less than 15 percent, this data may be substituted with values of one half of the reported method detection limit (MDL/2) or one half the method's practical quantitation limit (PQL/2). If more than 15 percent of the data are NDs, then non-parametric test procedures are recommended for evaluating the monitoring data. The non-parametric methods will treat the NDs values as ties; this procedure is less sensitive to error than parametric methods.

For NDs which represent greater than 50 percent of the data set, the test of proportions should be used. The test of proportions identifies the number of detections in a compliance well, relative to the background well, which is an indication of a statistically significant increase. These and other methods of handling nondetection data are discussed in the USEPA Unified Guidance.

VII. Comparison to Regulatory Limits

If groundwater data must be evaluated to determine whether a constituent has exceeded a regulatory limit (e.g., an MCL), a confidence interval approach based on the distribution of the data should be used. The confidence interval is designed to contain the true mean of the data with a specified level of confidence (generally 90 percent at the lower limit). The lower limit is then compared to the regulatory value and if the lower limit is larger, it is considered evidence that the regulatory level has been exceeded.

The data or log transformations of the raw data must meet normality assumptions when the student t-distribution is used. When nonparametric methods are employed, a minimum of seven independent data values must be available from a sampling event that is not affected by seasonality. Other distributions may be used to create the confidence interval so long as the data follows the assumed distribution. For further information, refer to the USEPA Unified Guidance.

VIII. Establishing Background Levels

High quality background data is the single most important key to a successful statistical groundwater monitoring program, especially for detection monitoring. All of the statistical tests listed in the Resource Conservation Recovery Act (RCRA) regulations are predicated on having appropriate and representative background measurements. The most important quality of background is that it reflects the historical conditions unaffected by the activities it is designed to be compared to.

In general, background refers to substances or locations that are not influenced by any releases from a site. The solid waste management rules require a groundwater monitoring system to have a minimum of one upgradient well and two downgradient monitoring wells for most solid waste facilities. For CCR facilities, a minimum of one upgradient well and three downgradient monitoring wells are required. In this case the upgradient well is considered to be the background well because it cannot be affected by the solid waste unit. But background may also refer to historic or baseline conditions within an individual well, such as performing intra-well time series plots or control charts.

To establish background at CCR facilities, NDAC Subsection 33.1-20-08-06(4) requires that a minimum of eight independent samples for each well must be collected and analyzed during the first six months of sampling for the detection monitoring program. NDAC Section 33.1-20-13-02 requires a minimum of four independent samples from each well to establish background for other solid waste facilities. The USEPA Unified Guidance recommends that a minimum of at least 8 to 10 independent background observations be collected before running most statistical tests. Although still a small sample size by statistical standards, these levels allow for minimally acceptable estimates of variability and evaluation of trend and goodness-of fit. However, this recommendation should be considered a temporary minimum until additional sampling can be conducted and the background sample size enlarged.

Due both to the complex behavior of groundwater and the need for sufficiently large sample sizes, background once obtained should not be regarded as a single fixed quantity. Naturally occurring background concentrations can change because of long term changes such as increased or decreased groundwater recharge, or other activities in the vicinity of a solid waste facility that are not related to leakage from the facility. Background should be periodically reviewed and revised as necessary throughout the life of the facility. There are no firm rules on how often to update background data. The general principle is that updating should occur when enough new measurements have been collected to allow a two-sample statistical comparison between the existing background data and a potential set of newer data. With semi-annual sampling, updating could be performed as often as every 2-3 years. Updating should generally not occur more frequently, since adding a new observation to background every one or two sampling rounds does not allow a statistical evaluation of whether the background mean is stationary over time.

Another practical aspect is that when background is updated, all statistical background limits (e.g., prediction and control limits) need to be recomputed to account for the revised background sample. At complex sites, updating the limits at each well and constituent on the monitoring list may require substantial effort. This includes resetting the cumulative sum [CUSUM] portions of control charts to zero after re-calculating the control limits and prior to additional testing against those limits. Too-frequent updating could thereby reduce the efficacy of control chart tests. However, it is recommended that background be reviewed for possible changes at least each time the facility permit is renewed, usually every ten years.

IX. References

Aitchison, J., and J.A.C. Brown, (1969). "The Lognormal Distribution" Cambridge University Press; Cambridge.

Dixon, W.J. and F.J. Massey, Jr., (1969). "Introduction to Statistical Analysis;" 3rd Edition; McGraw-Hill Book Co.; New York, New York.

Gibbons, J.D., (1976). "Non-parametric Methods for Quantitative Analysis;" Holt, Rinehart, and Winston Publishing Co.; New York, New York.

Gilbert, R.O., (1987). “Statistical Methods for Environmental Pollution Monitoring;” Van Nostrand Reinhold Co.; New York, New York.

USEPA, (2009). "Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities - Unified Guidance;" EPA 530/R-09-007; U.S. EPA; Office of Solid Waste Management; Washington, D.C. https://archive.epa.gov/epawaste/hazard/web/pdf/unified-guid-toc.pdf