WYOMING DEPARTMENT OF ENVIRONMENTAL QUALITY

SOLID AND HAZARDOUS WASTE DIVISION

SOLID WASTE GUIDELINE #14

GROUNDWATER MONITORING STATISTICAL ANALYSIS

I. Introduction

In compliance with EPA Subtitle D monitoring regulations, Chapter 2, Section 6(b)(i) of the Wyoming solid waste rules specifies groundwater monitoring and analysis requirements for Type I sanitary landfills. Specifically, Type I landfills that are required to monitor groundwater must institute a department approved detection monitoring program by the applicable date as defined in the rules. Further, the operator must evaluate the data with an approved statistical procedure, and submit an interpretation of the monitoring data to the department. If it is determined that a significant difference in water quality between background and down gradient wells exists, the department may require the facility to institute an approved assessment monitoring program. Again, the assessment monitoring program requires an approved statistical analysis of the monitoring data to be conducted and the results submitted to the department. Based on the analytical results of the assessment monitoring program, the department may require the facility to implement corrective action remedies as specified in Chapter 2, Section 8.

The final determination on whether groundwater quality has been impacted due to facility operations is significantly dependent upon the results of statistical analysis. Since there are numerous statistical procedures and software available for operator use, the department is concerned statistical results could potentially be inconsistent and/or inaccurate. In order to minimize these potential problems, the facility's statistical results will be evaluated by comparison with the results of statistical tests conducted by the department using Intelligent Decision Technologies, Ltd's "Groundwater Statistical Analysis System" (GSAS™) software. The department has reviewed GSAS™ software capabilities and determined the statistical procedures meet the regulatory requirements for groundwater monitoring data analysis.

The primary purpose of this guideline is to provide Type I sanitary landfill operators the data format for submitting monitoring data to the department on magnetic media. This data format is GSAS™ software compatible which will eliminate time-consuming data format conversions for the department. The department plans to use the submitted data for input to GSAS™ to verify facility submitted statistical analyses or to conduct independent statistical analyses prior to receiving facility required analyses. In addition, this guideline provides information and recommendations on statistical test methods available for monitoring data analysis. Please note, although Type II landfills are not currently required to statistically evaluate their groundwater monitoring data, the department requests future monitoring data are submitted to the department as specified in this guideline.

II. Regulation - Statistical Test Methods For Monitoring Data Analysis

Chapter 2, Section 6(b)(i)(C) of the solid waste rules specifies the approved statistical analysis procedures available, the required performance standards and the requirement for the facility owner/operator to determine whether there is a significant increase over background values for specific constituents. This portion of the guideline covers **basic** statistical test requirements for detection monitoring programs applicable to most Type I landfills in Wyoming. However, since each landfill situation is unique, the department recommends facilities thoroughly investigate the applicability of the suggested tests. The EPA guidance documents, referenced at the end of this guideline, provide a thorough description of statistical test methods and procedures for evaluating groundwater monitoring data.

In general, the appropriate statistical test used to evaluate groundwater quality is dependent upon many variables such as the type of monitoring program, data distribution, the number of upgradient and down gradient wells, the sampling interval, the number of sampling events, the number of non-detects for a specific constituent in the data set, single or multiple well comparisons, site hydrogeology, fate and transport characteristics of potential contamination at the site, etc.

The department has chosen the following **general** variable characteristics as applicable to the current groundwater monitoring of most Type I landfills.

1. The facility is conducting a department approved detection monitoring program. The most appropriate comparison of data for this type of monitoring is between a hydraulically upgradient (background) well and hydraulically down gradient (compliance) well(s), referred to as inter-well comparisons.

2. There is one hydraulically upgradient well and one or more hydraulically down gradient wells.

3. The sampling interval is quarterly for the first year of monitoring, which provides four consecutive sampling events for the first statistical analysis. As subsequent sampling events are added to the data set, statistical tests other than those suggested below may become applicable.

4. The percentage of non-detects for each constituent in each well is variable, and must be considered in choosing the appropriate statistical method. In general, EPA recommendations are: 1) if non-detects are less than 15% of all samples, each non-detect should be replaced by ½ its detection limit [recommendation is ½ of the practical quantitation limit (PQL) and using, not replacing, estimated concentrations], and proceed with a parametric one-way analysis; 2) if non-detects are between 15% and 90% of all samples then proceed with a non-parametric one-way analysis of variance procedure; and 3) if non-detects are greater than 90% of all samples, use a Poisson Prediction or Tolerance Limit procedure.

5. The site hydrogeology is such that all groundwater samples are derived from a continuous aquifer.

The following assumptions must be addressed prior to choosing a statistical procedure or before proceeding with a statistical analysis.

1. Normality Assumption:

Most statistical tests are based on the assumption that the data are normally distributed. Unless there are extreme violations of the normality assumptions, approved statistical tests usually provide accurate results. Therefore, EPA suggests facilities begin with the assumption that the original data are normal, and proceed with a parametric statistical test. If the test results in an erroneous analysis, take the natural logarithm of the original data and repeat the test. If the test results are incorrect again, the data must be checked for normality. There are several mathematical tests to check normality which are described in the referenced EPA guidance documents. Please keep in mind that EPA considers normality tests of original data for small sample sizes (i.e., fewer than 20 to 30 observations) to be questionable, and suggests tests conducted on transformed data are more accurate. Although the EPA recommended normality tests are not detailed here, it is recommended that facilities use probability plots first as a general graphical test for normality. If a computational normality check is required, EPA recommends the Shapiro-Wilk Test of Normality for sample sizes less than or equal to 50 and the Shapiro-Francia Test of Normality for sample sizes greater than 50. Please note that if the data grossly violate the normality assumption, a non-parametric statistical analysis must be used.

2. Equality of Variance Assumption:

The assumption that well groups (i.e., background and compliance wells) have approximately equal variances is important when using a parametric analysis of variance (ANOVA) statistical procedure. Please note that a parametric analysis ANOVA also requires that data are normally distributed, and usually if data are not normally distributed they also do not have equal variances. However, if data are transformed to create a normal

distribution, the transformation will also likely produce an equal variance. The results of a parametric ANOVA procedure are somewhat affected when the smallest and largest group variances differ by a ratio of 4 and the effect becomes severe for ratios of 10 or greater. EPA guidance outlines several procedures to test the assumption of equal variance as summarized below.

The department recommends facilities use "box plots" (as described in the referenced EPA guidance document) as a quick method to check for the equality of variance assumption. Other more formal procedures include Levene's Test and Bartlett's Test. Please note that the Bartlett's Test may give erroneous results if the data are not approximately normally distributed.

Statistical Methods

This guideline only covers those available statistical methods which the department considers most appropriate for the monitoring programs and data parameters as described above. These methods include the parametric one-way and non-parametric one-way analysis of variance (ANOVA) procedures and the Poisson Prediction and Tolerance Limits methods. Please note, as defined in Chapter 2 of the solid waste rules, other tolerance or prediction interval procedures, control charts, or other statistical methods approved by the administrator may be used to analyze the facility's monitoring data. The test requirements and data assumptions applicable to ANOVA and poisson distribution procedures follows:

1. The parametric one-way ANOVA is a statistical procedure which compares mean constituent concentrations between wells or groups of wells [i.e., background well(s) and compliance well(s)] to determine whether significant differences in mean concentrations are present.

The assumptions and requirements for this ANOVA procedure are as follows:

a. The assumptions that the data are normally distributed and equally variant are not grossly violated.
b. There are two or more wells being compared.
c. If the data show seasonal variations, corrections for seasonality must be made. If the data cannot be corrected, this procedure cannot be used.
d. Each well should have at least 3 observations and the difference between the total number of samples and the number of wells should be greater than or equal to 5.
e. The percentage of non-detects is less than 15%.

Note: EPA recommends the following procedure when using this ANOVA procedure. Plot the data versus time to check for seasonality and proceed as described above in 1(c), replace non-detects with ½ of the method detection limit (MDL) or preferably the PQL, calculate the one-way ANOVA, and check the residual values (difference between observed values and ANOVA predicted values) for normal distribution and equal variance as described above. If the residuals are normally distributed and equally variant the analysis is complete. If the residuals are not normally distributed and/or not equally variant, then transform the original data, calculate the ANOVA, and again check the residuals. If the residuals of the transformed data are normally distributed and equally variant the analysis is complete. A non-parametric procedure must be used if the original or transformed data are not normally distributed or equally variant.

2. A non-parametric one-way ANOVA statistical procedure is used when the parametric ANOVA procedure cannot be used due to data that are not normally distributed and/or not equally variant, or the number of non-detects is greater than 15%. Although non-detects are easily handled with this procedure, the percentage of non-detects cannot exceed 90%. In addition, this procedure can be used in any situation that the parametric ANOVA can be used.

The two non-parametric one-way ANOVA procedures recommended by EPA are the Kruskal-Wallis Rank Sum test and the Wilcoxon Rank Sum test. Both procedures replace the original data by their ranks and adjust the data for tied observations (i.e., ranking non-detects). The Kruskal-Wallis test requires a minimum of three wells (i.e., one background and two compliance wells) with at least four samples per well. The

Wilcoxon test is used when only one compliance well is being compared to one or more background wells, and the test requires at least four samples from each well.

3. Poisson Prediction and Tolerance Limits are statistical procedures recommended by EPA which model detected samples when 90% or more of the data values are non-detect.

Please note, tests for normality and equal variance and statistical analysis procedures can be easily computed and results obtained with statistics software specifically written for groundwater data analysis. Statistics programs which are not specifically designed for groundwater data may not offer appropriate statistical procedures or may require significant knowledge of statistics to obtain correct results. If you choose to perform the statistical calculations by hand, or require more detailed information, please refer to the EPA guidance documents referenced at the end of this guideline.

III. "Groundwater Statistical Analysis System" (GSAS™) Software

The department chose the GSAS™ software based primarily on the software's statistical analysis capabilities and procedures specifically designed to evaluate Subtitle C (hazardous waste facilities) and Subtitle D (municipal landfills) groundwater monitoring data in accordance with EPA regulation and guidance. In addition, Intelligent Decision Technologies, Ltd. provided the software at no cost for WDEQ's use as a government agency. The department disclaims any mention of GSAS™ or Intelligent Decision Technologies in this guideline as a department recommendation to purchase or use GSAS™ for your facility's groundwater monitoring statistical analysis requirements.

Although facilities may purchase or use any statistical software package, the department recommends that facilities choose a software package that includes, at a minimum, the following capabilities. Adequate statistical software should allow evaluation of pre-analysis tests, including checking the number of observations, data distribution, seasonality and outliers. Statistical software should also include the capability to perform inter-well and intra-well comparisons under detection, compliance and corrective action monitoring, and compliance limit comparisons under compliance and corrective action monitoring.

IV. Magnetic Media Data Format Requirements

This section describes the GSAS data format requirements for facilities submitting monitoring data to the department on magnetic media. Generally, the data can be input to most spreadsheet programs and output to a text format file, however, the file structure must be identical to the format described below.

Attached Exhibits A and B are examples of the above described data file format. Exhibit A illustrates a data file created in a spreadsheet program (EXCEL™ was used in this example). The data file appears as it would when entering the data (i.e., in program format), and the "boxes" facilitate tracking of columns and rows. After data entry is completed, save the file in both program and text (.txt) file formats, and submit the (.txt) file to the department. If a spreadsheet program is not available, a data text file can be created with many types of software, such as Notepad, which is a Microsoft WINDOWS™ application. The data file illustrated in Exhibit B is an example of a text file created with Notepad. Please note, in this format each row is one line and each column is designated by the column entry followed by a space. Also, at the beginning of a row (i.e., prior to entering any characters) each column is designated by one space.

The data file structure described below is applicable to Exhibits A and B described above. Please note, items in **all** lines (i.e., rows) may be separated by any number of spaces and/or tabs (i.e., spaces and/or tabs between columns), however, Exhibit B exhibits starting columns designated by one space.

First Line:

This line consists of general site information in the following order: Facility Name, Facility ID, Monitoring Program (e.g., detection), and Permit ID. Each item must be enclosed in square brackets and cannot exceed 20 characters. The only required item is the Facility Name, however, empty brackets must be used for items not used.

Second Line:

This line consists only of constituent names, and the first constituent name resides in the sixth column. Following the constituent name is **one** space followed by the units in parentheses. Although constituent names may be any number of words and spaces between words (within reason), the number of characters (including spaces) should be limited to 16 if possible. Please note, parentheses can only be used in conjunction with units, and that all data values for a specific constituent must have a consistent unit.

Third Line:

This line consists of CAS numbers. The number must be in the same column as the constituent name it represents. If CAS numbers are not used, n/a must be entered.

Fourth Line to End of File Line:

These lines consist of the well information and groundwater data. Each well's data are entered as a data group in the following format.

The First Well's Data Group Begins on the Fourth Line of the File:

The **first column** of this line (fourth line) contains the well name. The well name cannot contain any spaces (use an underscore if needed, e.g., MW_1) and only 8 characters are allowed. The **second column** contains a one letter designation for that well's gradient (i.e., enter u for an upgradient or background well and d for a down gradient or compliance well).

The next line (or fifth line since this well group began on the fourth line of the file) of this well's data group begins in the **third column** and contains the date in 00/00/00 format. The **fourth and fifth columns** must contain n/a. The **remaining columns** contain constituent values as described below.

1. The constituent value is entered in the corresponding column of the constituent name.
2. The units for the constituent value must be the same as identified in the constituent name.
3. The constituent value must be an integer or floating point number, and cannot exceed six decimal places. For example, if the reported value is 0.0000234 mg/l, the value could be changed to 0.0234 ug/l. However, in order to meet requirement #2 above, all other values in mg/l units for this specific constituent must also be changed to ug/l units (as well as the units identified with the constituent name).
4. Non-detects should be entered as a less than sign (<) followed by the method detection limit (MDL) or the practical quantitation limit (PQL). For example, <0.05.
5. Enter n/a if no sample was taken for that date/well/constituent combination.

Prior to creating the first data file, the department recommends facilities contact the department to verify the specific program the facility plans to use to create the data file is compatible with GSAS™ requirements, or submit an "example" output file.

In order for the department to check duplicate samples and/or verify data, the department requests facilities submit a paper copy of the original lab report with the data disk.

Further information can also be obtained from the following Solid and Hazardous Waste Division offices. Comments and suggestions for improvements are always appreciated.

Casper : (307) 473-3450 / Cheyenne : (307) 777-7752 / Lander : (307) 332-6942

Signed

_____ _____

David A. Finley Date
Administrator
Solid and Hazardous Waste Division

References

Office of Solid Waste, Waste Management Division, U.S. Environmental Protection Agency, Statistical Analysis of Groundwater at RCRA Facilities - Draft Guidance, October 1988.

Office of Solid Waste, Permits and State Programs Division, U.S. Environmental Protection Agency, Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities - Addendum to Interim Final Guidance, June 1992.

Attachments

Exhibit A - EXCEL™ Data File Structure (Program Format)
Exhibit B - Microsoft WINDOWS™ Notepad Data File Structure