Appendix B

Statistical Analysis Methods

Appendix B

Statistical Analysis Methods

INTRODUCTION

This appendix summarizes the statistical methods used to analyze programmatic data presented in this report.

LIQUID EFFLUENT MONITORING PROGRAM

Data Pretreatment and Validation

Liquid Effluent Monitoring Program data are validated following validation procedures to determine the quality of the analytical results. After the quality of the data is determined, program personnel assess the usability of the data. Data entry is also verified to prevent using inaccurate data results due to entry errors.

Control Charts

The control chart is a statistical tool used primarily to study a continuous process. For the Liquid Effluent Monitoring Program, the concentrations of analytes in the wastewater streams are the continuous processes of interest. While the concentrations of the analytes of interest for a specific stream are known to vary over time, plotting the values on a control chart can help assess the data for changes that might indicate a loss of process control or an unplanned release.

For each stream currently monitored, control charts are generated for each nonvolatile organic compound/nonradiological analyte with sufficient historical data to establish control limits. Available historical data from 1986 forward are used to generate the control limits. Current-year data are charted with the control limits to assess possible changes from historical stream characteristics. Currently, control limits are not calculated for radionuclides or volatile organic compounds due to the number of measurements below the detection limit and the lack of historical data prior to 1992.

By using control charts, it is assumed that the process is in control. Therefore, historical data are screened to exclude outliers and data from known periods when the effluent process changed. With the exception of pH, the concern is for unusually high concentrations. The control charts for these parameters are generated with a center line (based on the average of the historical data) and two upper control limits. The Level 1 upper control limits are calculated such that there is less than a 5% chance of exceeding the limit due to random fluctuations in the analyte concentration. For the Level 2 upper control limit, there is less than a 1% chance of exceeding the limit due to random fluctuations. Unusually low or high concentrations are both concerns for pH. Therefore, the pH control charts are generated with a lower and upper control limit. These limits are calculated such that there is less than a 1% chance that a concentration will fall outside either limit due to random fluctuations in the pH for the effluent.

Current year concentrations that exceed the Level 2 control limit (or either the upper or lower limit for pH) fall outside what is expected based on historical stream characteristics, but do not necessarily indicate an adverse environmental consequence. Instances where monitoring data exceed the Level 2 control limit (or either limit for pH) are reviewed to determine if a significant change occurred in the effluent stream or to determine if there are possible adverse environmental consequences. In most cases, no concern is identified. When the change is substantial and environmental or regulatory issues are identified, appropriate followup action is taken.

ENVIRONMENTAL SURVEILLANCES

Data Pretreatment

Before statistical analyses, data are screened to identify gross data errors, such as transcription errors, missing values, and out-of-range data points that do not meet other specific criteria, and to eliminate data from instruments that do not meet the minimum required operating characteristics as specified in the data quality objectives. After the initial screening, the data are screened for outliers. Graphical techniques, such as probability plots, stem and leaf plots, box plots, and other exploratory data analysis techniques, are the primary tools used for detecting potential data outliers. In cases where outliers are traceable to a specific error, a corrected value may be used to replace the outlier. If no correction is possible, then the point may be deleted from the data set. However, outliers with unattributable causes are rarely eliminated from data sets. Such outliers may be truly accurate data measurements indicative of unusual but important phenomena. Typically, two sets of analyses are performed, one with and one without the outlying data, and the two results are compared.

Trend Analyses

To visually evaluate long-term trends, cumulative data are presented graphically. For waste management surveillance gross alpha and gross beta air data, concentration data for specific locations are plotted over the year of interest.

For thermoluminescent dosimeter (TLD) data, cumulative six-month exposure data from specific locations, with background data (or distant community), are plotted over time. All historical data are smoothed and plotted on a linear scale to reveal the trend over time.

Comparisons Between Groupings

Penetrating Radiation Data from Thermoluminescent Dosimeters

Differences in yearly TLD exposures, either seasonally or by facility location, are analyzed using the nonparametric Kruskal-Wallis test for differences in medians. Nonparametric analyses are performed because the data are not expected to follow a normal distribution. Changes among groups are considered to be statistically significant if the p-value, associated with the null hypothesis, is less than 0.05. The null hypothesis is that the different samples in the groupings were from the same distribution or from distributions with the same median.

The statistical significance of changes in median exposures from the previous year to the current year is determined by facility. Facility groupings consist of background (or distant community) exposures, as well as individual waste management locations. Since the TLDs are changed every six months, the significance of the differences in the median seasonal exposure (either spring or fall) is also of interest.

Box and whisker plots graphically display the differences in median exposures between groups (either by facility or season). For each grouping, the median exposures of all the data is shown on the box and whisker plots, along with a box indicating the 25–75 percentile range based on all the data. The whiskers on the plots indicate the (nonoutlier) minimum and maximum exposures within each grouping.

For the box and whisker plots, the word “outlier” applies to those data values that are either greater than or less than 1.5 times the range of the box. This type of graph is used because it visually depicts differences in the medians of the groupings; therefore, the outliers are not shown since the scale required to show them would mask most of the visual differences in the medians calculation. Even though the outliers are not shown on the box and whisker plots, they are included in the calculation of the medians.

Airborne (Gross Alpha and Gross Beta) Data

Differences in year-to-year median concentrations for facility groupings of airborne data are also analyzed using the Kruskal-Wallis test for differences in medians. Data from the current year are grouped by facility for each contaminant and monitor type (that is, gross alpha or gross beta and PM_{10} or suspended particulate monitor). Differences in groupings are also graphically displayed using the box and whisker plots discussed above.