

TITLE: Identification of Outliers for Background Calculation

DATE EFFECTIVE: March 2009

HISTORY: New addition to the Technical Guidance Compendium

KEYWORDS: Outliers, background levels, sample population, sample point, chemicals of concern, Dixon's Test, Rosner's test, ProUCL, standard deviation, mean.

RULES: OAC 3745-300-01; 3745-300-07(H)

QUESTION: How are outliers calculated for background determination?

ANSWER: Background levels, are defined in OAC 3745-300-01 as "conditions at a property and areas surrounding a property that are unaffected by any current or past activities involving the treatment, storage, or disposal of hazardous substances or petroleum. Background levels include naturally occurring substances.

OAC 3745-300-07(H) allows background levels to be used as the applicable standard in lieu of a risk-based standard. Representative background sampling should be conducted in accordance with OAC 3745-300-07(H) to meet property-specific data quality objectives (DQOs). Sampling and Statistical approaches used to determine background levels for environmental media must be conducted in accordance with OAC 3745-300-07(H). For soil, the CP may use any statistically valid methodology to determine if there is a statistically significant difference between background samples and samples collected from identified areas. The CP may refer to "Guidance for Comparing Background and Chemical Concentrations in Soil for CERCLA sites" and "Statistical Methods for Evaluating the Attainment of Cleanup Standards" for guidance on calculating background levels.

Alternatively, a statistical method presented in OAC 3745-300-07(H)(1)(d) may be applied to establish background levels for soils. For calculating groundwater background levels, see OAC 3745-300-07(H)(3).

Prior to calculating the background level(s), the background dataset should be evaluated to determine if the dataset

contains outliers. The term outlier refers to a measured value that is numerically distant from, or not considered representative of the values within a particular data set. A “quick look” for outliers can be accomplished by plotting the data (normally or log normally) to see if the data points form a reasonably straight line when plotted on normal or lognormal axes. If the plot suggests the presence of one or more outliers, the dataset should be tested for outliers to determine whether there is statistical evidence that an observation appears extreme (whether upper bound or lower bound) and does not fit the distribution of the rest of the data. If a suspect observation is identified as an outlier, then steps need to be taken to determine whether it is the result of an error or a valid extreme observation. It is recommended that the CP select a method that does not incorporate the outlier itself in the calculations to determine outliers.

Please note that every data point that is above the maximum allowable limit is not necessarily an outlier. Conducting an outlier test on a data set *prior* to calculating a background value will provide a statistically robust background value.

Note that for soils, the CP/ Volunteer must collect a minimum of eight background soil samples within each zone or soil horizon for a representative background dataset. Any outlier observations removed from the dataset must be replaced with another observation, if needed, to maintain the minimum number of background samples required by the rules.

One approach to calculating outliers in the background dataset for soil, sediment and ground water is to use U.S. EPA’s ProUCL program which incorporates the statistics of Dixon and Rosner. Note is made that Dixon’s Test can be used for a single outlier in a data set of 25 or less samples, and Rosner’s Test can be used for up to five separate outliers within larger data sets. Both the Dixon and Rosner Tests can be run with the inclusion of the suspected outlier within the data set and therefore their results will not be dependent upon it. If the ProUCL program is not used, Dixon’s and Rosner’s tests are still acceptable for determining outliers in any data set. The ProUCL program can be downloaded from U.S. EPA’s website at <http://www.epa.gov/esd/tsc/download.htm> .

Outliers in ground water should be considered on a site specific or case by case basis, because the statistical tests for them are different than for outliers in soil. If the CP is concerned that concentrations of COCs in ground water are outliers, then the CP should contact the OEPA and request guidance for determining whether the concentration of the particular COC is an outlier.

The US EPA's Interim Final Guidance document: "Statistical Analysis of Ground Water Monitoring Data at RCRA Facilities presents ASTM E178-75 method for determining outliers and is another option to evaluate outliers in ground water if ProUCL is not used. Chapter 8 of the US EPA's Draft Unified Guidance document suggests that the previously referenced Dixon's and Rosner's Tests can be also be used for evaluation of outliers within ground water samples.

SUMMARY:

Calculating a background value must consider variations in data that may be outliers in a dataset. Various statistical tests can be used to support elimination of outliers, depending whether they are in soil or water, however the CP is encouraged to use the ProUCL program.

**OHIO EPA
CONTACT:**

For questions concerning this issue, please contact the VAP central office at: (614) 644-2924.