

Data Review and Exclusion Guidance For Determining Reasonable Potential and Deriving Water Quality-Based Effluent Limits

The Data Review and Exclusion Guidance (guidance) provides guidance to Missouri Department of Natural Resources' Water Protection Program, Clean Water Operating Permit Section permit writers, to assist in the determining when it is admissible to exclude data before determining reasonable potential or deriving effluent limits. The use of this guidance is specific to permit writers and not for either Compliance and Enforcement Section or Watershed Protection Section purposes.

Background

The purpose of the federal National Pollutant Discharge Elimination System is to address water pollution by regulating point sources that discharge pollutants to waters of the United States. As part of the permitting process, permit writers determine if specific parameters have reasonable potential to cause or contribute to an excursion above Missouri's Water Quality Standard per 40 CFR 122.44(d)(1)(i). If reasonable potential exists for a discharger to cause or contribute to an excursion of Missouri's Water Quality Standards, then permit writers must establish an effluent limit for that specific parameter, per 40 CFR 122.44(d)(1)(iii).

When determining reasonable potential, permit writers shall use five (5) years of a facility's Discharge Monitoring Reports (DMRs) data and data values from their application. Because data utilized for reasonable potential determination and effluent limit derivation has an important impact on water quality decisions, the data must be non-erroneous and representative of normal operations. Normal operations can include, but are not limited to, treatment adjustments, periods of optimization, and other common adjustments made by operators to address increased plant loading and other common operating activities.

Outliers

An outlier is a data point or points that lies outside of a expected statistical range of values in the dataset. Ignoring an outlier can skew the data or cause a problem to go unnoticed. Sometimes outliers show a significant error such as a sampling error. In other circumstances, the outlier can be real data, that is, a non-erroneous and representative data point, so the decision to exclude this data point is not a simple matter. The determination of an outlier only marks the data point as a candidate for exclusion rather than an automatic exclusion.

Justifiable Exclusions Based on Data Error or Non-representative Conditions

The justifiable exclusions listed below describe when data exclusion is allowable. These are categories to help permit writers identify circumstances where data removal is warranted; they are not final determinations for specific cases. The general rule is that if data is erroneous or does not represent normal operations, the data is subject to removal. However, data from normal operations is not eligible for a justifiable exclusion.

Each justifiable exclusion determination results from unique circumstances; therefore, the Department will consider all determinations on a case-by-case basis. Permit writers are to assist the permittee in determining if data is subject to exclusion or not.

It is primarily the permittee's responsibility to notify the permit writer that data may be subject to the one or more of the below justifiable exclusions and provide supporting data. However, if a permit writer notices a questionable data point, the permit writer may contact and work with the permittee to determine if the data point is subject to exclusion.

The justifiable exclusions are as follows:

- **Quality Assurance/Quality Control (QA/QC) Exclusion:** Data resulting from incorrect sampling, preservation, handling, testing, or analysis of the sample.
- **Data Entry Error Exclusion:** Data recorded incorrectly in the Department's DMR system. It is important to replace the incorrect data with real data, if available.
- **Industrial Contributor Exclusion:** Data from an industrial source that is no longer discharging to the treatment facility and contributing the pollutant of concern. The permit writer will evaluate the data to determine data exclusion. Things to consider during the determination include, but are not limited to, items such as residual strength, period of time since the facility received the discharge, and additional industrial sources with the same pollutant.
- **Changes in Technology Exclusion:** If the treatment system was subject to a change (e.g., upgrade or major modification), that affects the treatment of a parameter or parameters, the permit writer can exclude the data prior to the treatment change. However, if the change in treatment does not address a given parameter(s) then the permit writer should not excluded the data for that parameter based solely on the change in treatment. In the event that the treatment system's actual flow increased as part of the change, the permit writer can review the parameter data to determine if treatment or dilution is causing a noticeable difference in the parameter(s). Permittees should notify the permit writer of the dates of major changes so previous unrepresentative data may be excluded; the notification requirement is in accordance with 40 CFR 122.21(j) or 40 CFR 122.21(g) for industrial facilities.
- **Force Majeure Exclusion:** Flood, earthquake, and other natural disasters beyond the control of the technology or the operator.
- **Equipment Failure Exclusion:** Treatment/operational equipment failure resulting in above or below normal operations reflected in the data. Determinations for exclusion will include the permittee's responsibility of proper operations and maintenance per 40 CFR 122.41(e).
- **Start-Up Exclusion:** When a new treatment system begins operation or when an existing treatment system undergoes significant modifications (e.g., lagoon to mechanical plant), the DMR data for a period of time after the start-up may not depict the typical operations of the treatment facility.
- **Vandalism Exclusions:** Data results affected by vandalism outside of the permittee's control.
- **Slug Loading Exclusion:** The permit writer may exclude data resulting from a slug load depending on the permittee's prevention practices.

Statistical Tests

The permit writers accept DMR and application data submitted by the permittee as representative, and do not conduct statistical tests for the determination of outliers. If the permittee has determined that submitted data has quality issues, identified through statistical outlier test, then the permittee should notify the permit writer. Notification by itself will not result in the automatic removal of data without appropriate supporting information.

Below are the suggested statistical tests that a permittee can utilize to determine outliers based on normal distribution and non-symmetric distributions; however, this list is not exhaustive. If the permittee or consultant on their behalf submits an alternative outlier statistical test, the Department will need to review the test to determine appropriateness.

Normal Distribution Data	Skewed Data
Grubb's Test	Modified Z-Score
Tietjen-Moore test	Tukey's Method
Generalized Extreme Studentized Deviate Test or ESD Test	MADe Method
Interquartile Rule for Outliers or IQR Test	Median Rule
Normal Distribution: 2 Standard Deviation (Mean \pm 2SD)	Adjusted Boxplot
Z-Score	
Modified Z-Score	
Tukey's Method	
MADe Method	
Median Rule	
Adjusted Boxplot	

Standard Deviations

Standard deviation is a statistic that measures the dispersion or rather spread of data relative to its mean. Data with lower standard deviations are closer to the mean and data with higher standard deviations are farther from the mean. Standard deviation can also help in determining if data is representative of normal operations.

The empirical rule, also known as the three (3) sigma (σ) rule, states that nearly all the data (99.7%) will fall within three standard deviations of the mean. For non-normal distributed data datasets, the 3σ rule establishes that at least 88.8% of values will fall within three standard deviations. Thus, using this rule, the Department is reasonably certain that a significant majority of the data will fall within three standard deviations. Therefore, the Department considers any value at or above three standard deviations to not represent normal operations.

Data values that are three standard deviations from the mean are subject to exclusion. The cause of the value will assist in the determination if data removal should occur prior or after the determination of reasonable potential. Depending on the permittee's supporting information, data

exclusion may occur before determining reasonable potential. Regardless if supporting information provided, permit writer will exclude data above three standard deviations before effluent derivation.

Documentation in the Factsheet and Supporting Information

If the permittee request an exclusion, permit writers must explain, in the permit's factsheet and response correspondence, the decision to exclude, or not exclude. Permit writers may also review previous data to assist, if needed, to determine if exclusion is appropriate.

The permittee must provide supporting documentation justifying any request for exclusion of data and should state which category of justifiable exclusion applies. For example, if data is erroneous because of sample contamination, the permittee must provide supporting data, to include items such as duplicate testing results, field blanks, lab replicates, etc., in support of the contaminated sample claim. Another example is if the permittee indicates that the outlier is due to vandalism, then providing a police report would meet the supporting information requirement.