



Fact Sheet

Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities—Unified Guidance

Features of the Unified Guidance

What's new in the guidance? The March 2009 version of the Unified Guidance represents more than a decade of input from EPA Regions, states, statisticians working with groundwater monitoring, and results of a formal peer review. While the RCRA regulatory programs have been established for some time, existing guidance does not fully cover newer methods and experience gained in implementing the program. Major features include:

- Updated guidance for RCRA Subtitles C & D groundwater monitoring regulations covering all specified tests and performance criteria
- A suggested systematic detection monitoring framework to balance false positive errors and power in light of multiple comparisons
- Newer statistical methods for prediction limits, outlier, normality, autocorrelation and non-detect data diagnostic evaluations, and expanded use of non-parametric test methods
- Use of trend testing when stationarity assumptions cannot be met
- Expanded single-sample tests for compliance and corrective action monitoring, considering false positive errors and power

Organization. The guidance is laid out in four parts, with extensive Appendix statistical tables to support individual test methods:

- Part I identifies the key RCRA regulatory provisions and general recommendations for implementing these rules. It addresses issues of **statistical design**: factors such as developing and updating background data and strategies for constructing an effective statistical monitoring program.
- Part II covers diagnostic evaluations for checking key assumptions—outliers, normality, autocorrelation, non-detect data, spatial and temporal dependence. Useful exploratory techniques and tests are provided.
- Part III presents formal testing procedures for detection monitoring, covering both 40 CFR Parts 265, 264, and 258 requirements.
- Part IV is devoted to compliance and corrective action formal tests. Strategies are provided for a range of conditions including parametric and non-parametric alternatives.

What is the Unified Guidance?

This latest version of **Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities** is termed the Unified Guidance, since it integrates and supersedes two guidance documents of the same title released in 1989 and 1992. It resolves certain problems in earlier guidance while providing newer statistical methods and strategies developed in the mid-1990's and later. The guidance applies to both RCRA Subtitle C and D regulations. The focus is on RCRA hazardous and solid waste facility regulatory requirements, although the general statistical guidance is useful in other regulatory monitoring applications.

The guidance contains a compilation of statistical methods recommended for groundwater monitoring at RCRA and other facilities. It provides comprehensive strategies for designing the statistical aspects of facility detection, compliance, or corrective action monitoring systems. Interpretations are suggested for key statistical provisions of the RCRA groundwater monitoring regulations.

How was this guidance developed?

In the mid-1990's, the EPA Office of Solid Waste convened a task group consisting of state and EPA personnel, industry representatives, and statisticians closely involved with groundwater monitoring issues. The goal was to develop more current and relevant RCRA statistical guidance. Following a number of preliminary drafts, a full version was circulated in 2004 to interested state regulatory personnel for their comments, as well as to three expert peer reviewers in 2005. The various drafts were produced by Science Applications International Corporation (SAIC), using the technical expertise of statistician Dr. Kirk Cameron (MacStat Consulting Ltd). The Unified Guidance has been substantially modified and expanded to address the issues raised by commenters.

Who are potential users of this guidance?

The guidance is aimed at the informed professional working in the groundwater monitoring field, assuming a limited background in statistics. The primary users are expected to be:

- Owners, operators, and personnel at Subtitle C hazardous waste or Subtitle D solid waste facilities
- State and EPA regulatory personnel concerned with permits, enforcement and compliance at these facilities
- Consultants and statisticians providing technical assistance to regulated facilities; and
- Other ground water and regulatory monitoring program personnel such as in the CERCLA program.

Features of the Unified Guidance

Part I-- Introductory Framework

- Regulatory Issues
 - Hypothesis testing frameworks
 - Sampling requirements
 - Limitations of certain tests like ANOVA
- The groundwater monitoring context
- Basic statistical concepts
- The nature of hypothesis testing
- Establishing and updating background data

- Detection Monitoring Design
 - Control of false positive errors with multiple comparisons
 - Sitewide False Positive Error Rate [SWFPR] application
 - Minimum power reference criteria
 - Using multiple test methods
 - Effect size power evaluation
 - Appropriate tests including trend analysis
- Compliance/Corrective Action Monitoring Design
 - Use of single sample tests against a fixed standard
 - Hypothesis framework
 - Centrality versus upper percentile parameters
 - Test types (parametric vs. non-parametric, trends)
 - Testing Against a Background Standard

Part II-- Diagnostic Evaluation and Testing

- Exploratory data tools
- Goodness-of-fit testing
 - Importance of the normal distribution
 - Other normalizing transformations (logarithmic, ladder-of-powers)
- Outliers
- Equality of Variance
- Managing Non-Detect Data
- Spatial Dependence
- Types of Temporal Dependence
 - autocorrelation, trends, seasonality, etc.

Part III-- Detection Monitoring Tests

- Coverage of all regulatory tests
 - t-tests, ANOVA, control charts, prediction and tolerance limits
- Parametric versus non-parametric methods
- Tests when non-detect data are present
- Use of trend analyses
- Emphasis on prediction limits for systematic design

Part IV-- Compliance/Corrective Action Tests

- Test of means versus upper percentiles
- Control of false positive errors and power
- Fixed standards vs. background limits

What legal limitations does this guidance impose?

EPA makes it clear at the outset of the document that this present work is **guidance only**, and does not confer any legal requirements or obligations on regulated entities or regulatory programs. While it is necessary to make interpretations of regulatory language to apply statistical measures, those found in the guidance are only suggested. Other approaches and statistical methods can work equally well or better in specific instances. As a practical matter, it is recognized that states may choose to adopt requirements similar to guidance recommendations. While we believe that the document offers reasonable current guidance, experience and statistical applications in this field are continually evolving.

What regulations and issues are covered?

The guidance covers the statistical aspects of groundwater monitoring regulations for 40 CFR Parts 265, 264, and 258. These include monitoring under Subtitle C interim status and RCRA permits, as well as for Subtitle D solid waste facilities. These rules span a considerable period of time from 1980 forward, with significant modifications to the Part 264 regulations in 1988 and 2006. Key portions of regulatory language pertaining to groundwater monitoring and statistical testing are provided in the guidance. These include the specified test procedures, performance criteria, sampling requirements, and identification of relevant groundwater protection standards.

Basic statistical interpretations include identifying the appropriate hypothesis testing frameworks, meeting performance criteria, the application of certain sampling data requirements, and the use and limitation of designated tests. For some applications, the regulations do not explicitly identify appropriate test methods; the Unified Guidance makes reasonable judgments as to the more appropriate procedures. One particular issue stressed throughout the guidance is the need to utilize statistically **independent** data as identified in 1988 and later RCRA regulatory language. Certain regulatory restrictions also dictate the appropriate responses for RCRA applications, but may not be limiting in other monitoring situations.

How is this document organized?

The guidance follows a logical progression from simple and general discussions to more detailed coverage of specific test methods. After presenting the regulatory context in Part I, a chapter is devoted to basic statistical concepts. These include the assumptions found in the RCRA performance criteria but are more broadly extended to include other standard statistical factors. Terms such as independence, statistical significance, stationarity, random sampling, spatial and temporal dependence, normality, equality of variance, outliers and non-detect data are defined and explained. The overall groundwater monitoring context is presented, with special emphasis on hypothesis testing and the related false positive and negative errors. A separate chapter discusses developing, assessing and updating background data.

General design considerations are provided for developing a detection monitoring system. The guidance provides a systematic approach to integrating false positive errors and power in a site design. We specifically recommend a 10% Site-Wide False Positive Rate [SWFPR] partitioned among the total number of tests per year. EPA Reference Power Curves [ERPC] are provided as minimum criteria for sufficient statistical power, used to gauge the effectiveness of particular detection monitoring tests.

Design of compliance or corrective action monitoring systems follows. Because most groundwater protection standards [GWPS] are in the form of fixed, risk- or health-based limits, the design differs along with the appropriate types of statistical tests. Unlike highly site-specific detection programs, key decisions need to be made by regulatory agencies. These include the appropriate type of parameter for comparison to the GWPS, false positive and negative error rates, and the form of hypothesis testing. The use of a background GWPS is also discussed.

Following a summary chapter of recommended methods, detailed consideration of diagnostic evaluations and testing of data are provided in Part II. These include general exploratory techniques such as box plots or probability plots, testing for goodness-of-fit, outliers, non-detect data, equality of variance, spatial and temporal dependence. If assumptions critical to statistical tests are not met, the guidance suggests potential data adjustments for these situations.

Part III provides the specific detection monitoring tests found in the RCRA regulations. Each test is discussed in overall terms including necessary assumptions, followed by a detailed procedure and example. All formal tests in the guidance follow this same approach.

Part IV contains detailed methods for compliance and assessment monitoring using confidence intervals. Consideration is given to the design aspects presented earlier, including the parameter choice and hypothesis framework. A discussion of cumulative false positive errors and power is provided. Depending on whether compliance or corrective action monitoring is involved, false positive error and power criteria can vary based on different perspectives of the regulated entity and agency. The guidance offers recommendations which place priority on EPA and state regulatory needs to enhance protection of public health and the environment.

The appendices contain references, a glossary and index, as well as extensive tables for specific test methods which span the range of conditions likely to occur at regulated facilities.

Why is it recommended to use the SWFPR and ERPC in detection monitoring design?

These criteria stem from problems historically experienced at facilities conducting multiple statistical tests for a wide range of monitoring constituents at numerous compliance wells. This is the classic **multiple comparisons problem**. When many tests are conducted at a fixed error rate, the chances of one or more false positive errors (a condition when one concludes that a release has occurred when there is in fact none) can become unreasonably high. A second and very important consideration is that statistical tests must have sufficient ability (or power) to detect such a release when it occurs.

Within the limits of the RCRA regulations, certain opportunities were afforded to control this potentially high rate of false positive error. This is especially true if prediction limits are used as tests, although two other identified methods—control charts and tolerance limits—can be similarly designed. By maintaining a consistent overall annual error rate, all regulated facilities will be afforded the same risk.

Based on earlier work by EPA and others, prediction limit tests typical of the RCRA groundwater monitoring context were identified as a minimally acceptable criterion for power to detect real releases to groundwater. While a relative measure, it can be applied universally to all detection monitoring tests. The March 2009 Unified Guidance extends this approach to consider the cumulative power of tests, based on the number of annual evaluations per year. It provides a common framework for considering both cumulative false positive errors and power.

The guidance also discusses effect size power as an alternative to the relative power criteria. This approach requires a regulatory agency determination of a specific increase of concern. At present, there are few if any such criteria established. This approach may find use in specific applications discussed in the guidance.

While the SWFPR and ERPC approaches are recommended for detection monitoring, the guidance reaches different conclusions for compliance and corrective action monitoring when fixed limits are used as standards. The situation is too uncertain and problematic to apply the same concepts, and other strategies are recommended.

Why is diagnostic testing important and when should it be used?

In addition to addressing the RCRA regulatory requirements for performance criteria, it is good statistical

practice to know one's data closely. Checking key assumptions is critical to proper performance of any statistical test. Misapplication can also generate results which do not follow the expected outcomes of a given test. Diagnostic testing is performed primarily during permit or remedial action plan development. Once a set of tests is selected for formal permit or remedial plan monitoring, diagnostic testing might only be periodically expected (e.g., for updating background data).

Many important statistical tests assume a **normal distribution**. **Goodness-of-fit** techniques for identifying a probable normal distribution are found in the guidance. In many situations, a transformation of data (e.g., logarithmic, square root) can result in approximately normal data. Other parametric distributions may work equally well or better in some situations, but the guidance generally focuses on the family of normal distributions. If no transformation is suitable, non-parametric test methods can be used.

Equality of variance is an additional assumption necessary for some tests. The guidance provides both exploratory measures and a formal statistical test.

Outliers, often very large values of dubious quality, can significantly weaken the ability of tests to perform as expected. The guidance offers two test methods for identifying outliers, and suggestions for when they might be removed, replaced or otherwise avoided.

Spatial variability is a very important consideration. If background monitoring constituent mean data vary by well, assumptions for certain detection monitoring tests like Analysis of Variance (ANOVA) will not be met. More importantly, it will generally be impossible to determine if mean well differences are due to existing background conditions or a true release. Parametric or non-parametric ANOVAs are recommended in the guidance as diagnostic tests to initially establish if prior spatial differences exist. The outcomes may vary with the types of constituents being monitored.

Several forms of **temporal variation** can occur. Temporal variation is some non-random pattern in data over time. It could include autocorrelation, seasonal variation, well-to-well constituent correlation, correlation among monitoring constituents in a well, and the presence of trends. Each of these types of temporal dependence requires somewhat different diagnostic testing and potential adjustments provided in the guidance.

Non-detect values are a common feature of many RCRA constituent data sets. Those containing multiple non-detect limits are of particular concern. The Unified Guidance provides a number of non-detect data adjustment procedures, including two fairly recent methods for multiple non-detect limits.

Which detection monitoring tests are recommended?

While the guidance covers all of the regulatory tests, there is a clear preference for prediction limits or control charts as detection monitoring tests. The guidance specifically recommends the Shewhart-CUSUM option when choosing control charts.

For interim status or facilities with few annual tests, variants of the Student-t or alternative non-parametric two-sample tests may be sufficient. Other facilities will need to apply tests which account for the multiple comparisons. Both because of the common presence of spatial variability and regulatory restrictions, neither parametric nor non-parametric ANOVA tests are likely to be used frequently. Tolerance limits are similar to prediction limits, but their usefulness in designing a systematic detection monitoring program is more limited. Prediction limits provide the greatest flexibility, and the guidance provides the most extensive details for this method. By careful use of repeat testing, prediction limits can minimize future sampling requirements and meet the SWFPR and ERPC criteria. Nine different parametric and six non-parametric variants are provided to address most monitoring situations.

Which compliance/corrective action monitoring tests are recommended?

The regulatory agency first determines the appropriate form of comparison to groundwater protection standards [GWPS]. The guidance offers a number of single-sample tests for centrality parameters such as the arithmetic mean, geometric mean, arithmetic mean of a lognormal distribution, and median tests. If the decision is that a maximum limit is appropriate, the guidance offers parametric or non-parametric upper percentiles as options. Confidence intervals around trend lines may be appropriate in some instances. Testing background GWPS can either use options provided here or those for detection monitoring.

Where can the public get more information about this guidance?

The guidance will be available on the EPA website: <http://www.epa.gov/epawaste/hazard/correctiveaction/resources/guidance/sitechar/gwstats/index.htm>. For further assistance, please contact Mike Gansecki, EPA Region 8 (email: gansecki.mike@epa.gov or by phone: (303-312-6150).