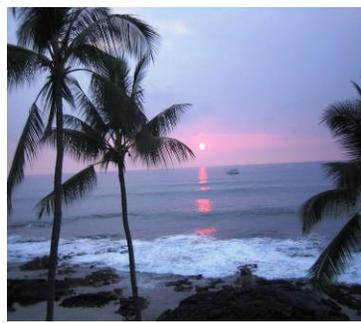
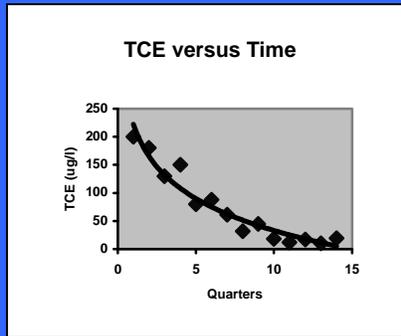


**NATURAL RESOURCES DEFENSE COUNCIL'S & POWDER RIVER BASIN  
RESOURCE COUNCIL'S PETITION FOR REVIEW**

**EXHIBIT 13**

# STATISTICAL ANALYSIS OF GROUNDWATER MONITORING DATA AT RCRA FACILITIES UNIFIED GUIDANCE MARCH 2009

*EPA 530/R-09-007*



**ENVIRONMENTAL PROTECTION AGENCY  
OFFICE OF RESOURCE CONSERVATION AND RECOVERY**



## TABLE OF CONTENTS

DISCLAIMER.....	i
ACKNOWLEDGMENTS.....	ii
EXECUTIVE SUMMARY.....	iii
TABLE OF CONTENTS.....	v

### PART I. STATISTICAL DESIGN AND PHILOSOPHY

#### CHAPTER 1. OBJECTIVES AND POTENTIAL USE OF THIS GUIDANCE

1.1 OBJECTIVES.....	1-1
1.2 APPLICABILITY TO OTHER ENVIRONMENTAL PROGRAMS.....	1-3

#### CHAPTER 2. REGULATORY OVERVIEW

2.1 REGULATORY SUMMARY.....	2-1
2.2 SPECIFIC REGULATORY FEATURES AND STATISTICAL ISSUES.....	2-6
2.2.1 Statistical Methods Identified under §264.97(h) and §258.53(g).....	2-6
2.2.2 Performance Standards under §264.97(i) and §258.53(h).....	2-7
2.2.3 Hypothesis Tests in Detection, Compliance and Corrective Action Monitoring.....	2-10
2.2.4 Sampling Frequency Requirements.....	2-10
2.2.5 Groundwater Protection Standards.....	2-12
2.3 UNIFIED GUIDANCE RECOMMENDATIONS.....	2-13
2.3.1 Interim Status Monitoring.....	2-13
2.3.2 Parts 264 and 258 Detection Monitoring Methods.....	2-14
2.3.3 Parts 264 and 258 Compliance/assessment Monitoring.....	2-15

#### CHAPTER 3. KEY STATISTICAL CONCEPTS

3.1 INTRODUCTION TO GROUNDWATER STATISTICS.....	3-2
3.2 COMMON STATISTICAL ASSUMPTIONS.....	3-4
3.2.1 Statistical Independence.....	3-4
3.2.2 Stationarity.....	3-5
3.2.3 Lack of Statistical Outliers.....	3-7
3.2.4 Normality.....	3-7
3.3 COMMON STATISTICAL MEASURES.....	3-9
3.4 HYPOTHESIS TESTING FRAMEWORK.....	3-12
3.5 ERRORS IN HYPOTHESIS TESTING.....	3-14
3.5.1 False Positives and Type I Errors.....	3-15
3.5.2 Sampling Distributions, Central Limit Theorem.....	3-16
3.5.3 False Negatives, Type II Errors, and Statistical Power.....	3-18
3.5.4 Balancing Type I and Type II Errors.....	3-22

#### CHAPTER 4. GROUNDWATER MONITORING PROGRAMS AND STATISTICAL ANALYSIS

4.1 THE GROUNDWATER MONITORING CONTEXT.....	4-1
4.2 RCRA GROUNDWATER MONITORING PROGRAMS.....	4-3

4.3 STATISTICAL SIGNIFICANCE IN GROUNDWATER TESTING..... 4-6

    4.3.1 *Statistical Factors*..... 4-8

    4.3.2 *Well System Design and Sampling Factors*..... 4-8

    4.3.3 *Hydrological Factors*..... 4-9

    4.3.4 *Geochemical Factors*..... 4-10

    4.3.5 *Analytical Factors*..... 4-10

    4.3.6 *Data or Analytic Errors*..... 4-11

CHAPTER 5. ESTABLISHING AND UPDATING BACKGROUND

5.1 IMPORTANCE OF BACKGROUND..... 5-1

    5.1.1 *Tracking Natural Groundwater Conditions*..... 5-2

5.2 ESTABLISHING AND REVIEWING BACKGROUND..... 5-2

    5.2.1 *Selecting Monitoring Constituents and Adequate Sample Sizes*..... 5-2

    5.2.2 *Basic Assumptions About Background*..... 5-4

    5.2.3 *Outliers in Background*..... 5-5

    5.2.4 *Impact of Spatial Variability*..... 5-6

    5.2.5 *Trends in Background*..... 5-7

    5.2.6 *Summary: Expanding Background Sample Sizes*..... 5-8

    5.2.7 *Review of Background*..... 5-10

5.3 UPDATING BACKGROUND..... 5-12

    5.3.1 *When to Update*..... 5-12

    5.3.2 *How to Update*..... 5-12

    5.3.3 *Impact of Retesting*..... 5-14

    5.3.4 *Updating When Trends are Apparent*..... 5-14

CHAPTER 6. DETECTION MONITORING PROGRAM DESIGN

6.1 INTRODUCTION..... 6-1

6.2 Elements of the Statistical Program Design..... 6-2

    6.2.1 *The Multiple Comparisons Problem*..... 6-2

    6.2.2 *Site-Wide False Positive Rates [SWFPR]*..... 6-7

    6.2.3 *Recommendations for Statistical Power*..... 6-13

    6.2.4 *Effect Sizes and Data-Based Power Curves*..... 6-18

    6.2.5 *Sites Using More Than One Statistical Method*..... 6-21

6.3 HOW KEY ASSUMPTIONS IMPACT STATISTICAL DESIGN..... 6-25

    6.3.1 *Statistical Independence*..... 6-25

    6.3.2 *Spatial Variation: Interwell vs. Intrawell Testing*..... 6-29

    6.3.3 *Outliers*..... 6-34

    6.3.4 *Non-Detects*..... 6-36

6.4 DESIGNING DETECTION MONITORING TESTS..... 6-38

    6.4.1 *T-Tests*..... 6-38

    6.4.2 *Analysis Of Variance [ANOVA]*..... 6-38

    6.4.3 *Trend Tests*..... 6-41

    6.4.4 *Statistical Intervals*..... 6-42

    6.4.5 *Control Charts*..... 6-46

6.5 SITE DESIGN EXAMPLES..... 6-46

CHAPTER 7. STRATEGIES FOR COMPLIANCE/ASSESSMENT MONITORING AND CORRECTIVE ACTION

7.1 INTRODUCTION..... 7-1

7.2 HYPOTHESIS TESTING STRUCTURES..... 7-3

7.3 GROUNDWATER PROTECTION STANDARDS..... 7-6

7.4 DESIGNING A STATISTICAL PROGRAM..... 7-9

    7.4.1 *False Positives and Statistical Power in Compliance/Assessment*..... 7-9

    7.4.2 *False Positives and Statistical Power In Corrective Action*..... 7-12

    7.4.3 *Recommended Strategies*..... 7-13

    7.4.4 *Accounting for Shifts and Trends*..... 7-14

    7.4.5 *Impact of Sample Variability, Non-Detects, And Non-Normal Data*..... 7-17

7.5 COMPARISONS TO BACKGROUND DATA..... 7-20

CHAPTER 8. SUMMARY OF RECOMMENDED METHODS

8.1 SELECTING THE RIGHT STATISTICAL METHOD..... 8-1

8.2 TABLE 8.1 INVENTORY OF RECOMMENDED METHODS..... 8-4

8.3 METHOD SUMMARIES..... 8-9

PART II. DIAGNOSTIC METHODS AND TESTING

CHAPTER 9. COMMON EXPLORATORY TOOLS

9.1 TIMES SERIES PLOTS..... 9-1

9.2 BOX PLOTS..... 9-5

9.3 HISTOGRAMS..... 9-8

9.4 SCATTER PLOTS..... 9-13

9.5 PROBABILITY PLOTS..... 9-16

CHAPTER 10. FITTING DISTRIBUTIONS

10.1 IMPORTANCE OF DISTRIBUTIONAL MODELS..... 10-1

10.2 TRANSFORMATIONS TO NORMALITY..... 10-3

10.3 USING THE NORMAL DISTRIBUTION AS A DEFAULT..... 10-5

10.4 COEFFICIENT OF VARIATION AND COEFFICIENT OF SKEWNESS..... 10-9

10.5 SHAPIRO-WILK AND SHAPIRO-FRANCÍA NORMALITY TESTS..... 10-13

    10.5.1 *Shapiro-Wilk Test ( $n \leq 50$ )*..... 10-13

    10.5.2 *Shapiro-Francis Test ( $n > 50$ )*..... 10-15

10.6 PROBABILITY PLOT CORRELATION COEFFICIENT..... 10-16

10.7 SHAPIRO-WILK MULTIPLE GROUP TEST OF NORMALITY..... 10-19

CHAPTER 11. TESTING EQUALITY OF VARIANCE

11.1 BOX PLOTS ..... 11-2

11.2 LEVENE'S TEST..... 11-4

11.3 MEAN-STANDARD DEVIATION SCATTER PLOT..... 11-8

CHAPTER 12. IDENTIFYING OUTLIERS

12.1 SCREENING WITH PROBABILITY PLOTS..... 12-1  
 12.2 SCREENING WITH BOX PLOTS..... 12-5  
 12.3 DIXON’S TEST..... 12-8  
 12.4 ROSNER’S TEST..... 12-10

CHAPTER 13. SPATIAL VARIABILITY

13.1 INTRODUCTION TO SPATIAL VARIATION..... 13-1  
 13.2 IDENTIFYING SPATIAL VARIABILITY..... 13-2  
     13.2.1 Side-by-Side Box Plots..... 13-2  
     13.2.2 One-Way Analysis of Variance for Spatial Variability..... 13-5  
 13.3 USING ANOVA TO IMPROVE PARAMETRIC INTRAWELL TESTS..... 13-8

CHAPTER 14. TEMPORAL VARIABILITY

14.1 TEMPORAL DEPENDENCE..... 14-1  
 14.2 IDENTIFYING TEMPORAL EFFECTS AND CORRELATION..... 14-3  
     14.2.1 Parallel Time Series Plots..... 14-3  
     14.2.2 One-Way Analysis of Variance for Temporal Effects..... 14-6  
     14.2.3 Sample Autocorrelation Function..... 14-12  
     14.2.4 Rank von Neumann Ratio Test..... 14-16  
 14.3 CORRECTING FOR TEMPORAL EFFECTS AND CORRELATION..... 14-18  
     14.3.1 Adjusting the Sampling Frequency and/or Test Method..... 14-18  
     14.3.2 Choosing a Sampling Interval Via Darcy’s Equation..... 14-19  
     14.3.3 Creating Adjusted, Stationary Measurements..... 14-28  
     14.3.4 Identifying Linear Trends Amidst Seasonality: Seasonal Mann-Kendall Test.... 14-37

CHAPTER 15. MANAGING NON-DETECT DATA

15.1 GENERAL CONSIDERATIONS FOR NON-DETECT DATA..... 15-1  
 15.2 IMPUTING NON-DETECT VALUES BY SIMPLE SUBSTITUTION..... 15-3  
 15.3 ESTIMATION BY KAPLAN-MEIER..... 15-7  
 15.4 ROBUST REGRESSION ON ORDER STATISTICS..... 15-13  
 15.5 OTHER METHODS FOR A SINGLE CENSORING LIMIT..... 15-21  
     15.5.1 Cohen’s Method..... 15-21  
     15.5.2 Parametric Regression on Order Statistics..... 15-23  
 15.6 USE OF THE 15% AND 50% NON-DETECT RULE..... 15-24

PART III. DETECTION MONITORING TESTS

CHAPTER 16. TWO-SAMPLE TESTS

16.1 PARAMETRIC T-TESTS..... 16-1  
     16.1.1 Pooled Variance T-Test..... 16-4  
     16.1.2 Welch’s T-Test..... 16-7  
     16.1.3 Welch’s T-Test and Lognormal Data..... 16-10  
 16.2 WILCOXON RANK-SUM TEST..... 16-14  
 16.3 TARONE-WARE TWO-SAMPLE TEST FOR CENSORED DATA..... 16-20

CHAPTER 17. ANOVA, TOLERANCE LIMITS, AND TREND TESTS

17.1 ANALYSIS OF VARIANCE [ANOVA]..... 17-1  
 17.1.1 One-Way Parametric F-Test..... 17-1  
 17.1.2 Kruskal-Wallis Test..... 17-9  
 17.2 TOLERANCE LIMITS..... 17-14  
 17.2.1 Parametric Tolerance Limits..... 17-15  
 17.2.2 Non-Parametric Tolerance Intervals..... 17-18  
 17.3 TREND TESTS..... 17-21  
 17.3.1 Linear Regression..... 17-23  
 17.3.2 Mann-Kendall Trend Test..... 17-30  
 17.3.3 Theil-Sen Trend Line..... 17-34

CHAPTER 18. PREDICTION LIMIT PRIMER

18.1 INTRODUCTION TO PREDICTION LIMITS..... 18-1  
 18.1.1 Basic Requirements for Prediction Limits..... 18-4  
 18.1.2 Prediction Limits With Censored Data..... 18-6  
 18.2 PARAMETRIC PREDICTION LIMITS..... 18-7  
 18.2.1 Prediction Limit for *m* Future Values..... 18-7  
 18.2.2 Prediction Limit for a Future Mean..... 18-11  
 18.3 NON-PARAMETRIC PREDICTION LIMITS..... 18-16  
 18.3.1 Prediction Limit for *m* Future Values..... 18-17  
 18.3.2 Prediction Limit for a Future Median..... 18-20

CHAPTER 19. PREDICTION LIMIT STRATEGIES WITH RETESTING

19.1 RETESTING STRATEGIES ..... 19-1  
 19.2 COMPUTING SITE-WIDE FALSE POSITIVE RATES [SWFPR]..... 19-4  
 19.2.1 Basic Subdivision Principle..... 19-7  
 19.3 PARAMETRIC PREDICTION LIMITS WITH RETESTING..... 19-11  
 19.3.1 Testing Individual Future Values..... 19-15  
 19.3.2 Testing Future Means..... 19-20  
 19.4 NON-PARAMETRIC PREDICTION LIMITS WITH RETESTING..... 19-26  
 19.4.1 Testing Individual Future Values..... 19-30  
 19.4.2 Testing Future Medians..... 19-31

CHAPTER 20. MULTIPLE COMPARISONS USING CONTROL CHARTS

20.1 INTRODUCTION TO CONTROL CHARTS..... 20-1  
 20.2 BASIC PROCEDURE..... 20-2  
 20.3 CONTROL CHART REQUIREMENTS AND ASSUMPTIONS..... 20-6  
 20.3.1 Statistical Independence and Stationarity..... 20-6  
 20.3.2 Sample Size, Updating Background..... 20-8  
 20.3.3 Normality and Non-Detect Data..... 20-9  
 20.4 CONTROL CHART PERFORMANCE CRITERIA..... 20-11  
 20.4.1 Control Charts with Multiple Comparisons..... 20-12  
 20.4.2 Retesting in Control Charts..... 20-14

PART IV. COMPLIANCE/ASSESSMENT AND CORRECTIVE ACTION TESTS

CHAPTER 21. CONFIDENCE INTERVALS

21.1 PARAMETRIC CONFIDENCE INTERVALS..... 21-1  
     21.1.1 Confidence Interval Around a Normal Mean..... 21-3  
     21.1.2 Confidence Interval Around a Lognormal Geometric Mean..... 21-5  
     21.1.3 Confidence Interval Around a Lognormal Arithmetic Mean..... 21-8  
     21.1.4 Confidence Interval Around an Upper Percentile..... 21-11  
 21.2 NON-PARAMETRIC CONFIDENCE INTERVALS..... 21-14  
 21.3 CONFIDENCE INTERVALS AROUND TREND LINES..... 21-23  
     21.3.1 Parametric Confidence Band Around Linear Regression..... 21-23  
     21.3.2 Non-Parametric Confidence Band Around a Theil-Sen Line..... 21-30

CHAPTER 22. COMPLIANCE/ASSESSMENT AND CORRECTIVE ACTION TESTS

22.1 CONFIDENCE INTERVAL TESTS FOR MEANS..... 22-1  
     22.1.1 Pre-Specifying Power In Compliance/Assessment..... 22-2  
     22.1.2 Pre-Specifying False Positive Rates in Corrective Action..... 22-9  
 22.2 CONFIDENCE INTERVAL TESTS FOR UPPER PERCENTILES..... 22-18  
     22.2.1 Upper Percentile Tests in Compliance/Assessment..... 22-19  
     22.2.2 Upper Percentile Tests in Corrective Action..... 22-20

APPENDICES<sup>1</sup>

A.1 REFERENCES  
 A.2 GLOSSARY  
 A.3 INDEX  
 B HISTORICAL NOTES  
 C.1 SPECIAL STUDY: NORMAL VS. LOGNORMAL PREDICTION LIMITS  
 C.2 CALCULATING STATISTICAL POWER  
 C.3 R SCRIPTS  
 D STATISTICAL TABLES

---

<sup>1</sup> The full table of contents for Appendices A through D is found at the beginning of the Appendices

## CHAPTER 5. ESTABLISHING AND UPDATING BACKGROUND

5.1	IMPORTANCE OF BACKGROUND .....	5-1
5.1.1	<i>Tracking Natural Groundwater Conditions</i> .....	5-2
5.2	ESTABLISHING AND REVIEWING BACKGROUND .....	5-2
5.2.1	<i>Selecting Monitoring Constituents and Adequate Sample Sizes</i> .....	5-2
5.2.2	<i>Basic Assumptions About Background</i> .....	5-4
5.2.3	<i>Outliers in Background</i> .....	5-5
5.2.4	<i>Impact of Spatial Variability</i> .....	5-6
5.2.5	<i>Trends in Background</i> .....	5-7
5.2.6	<i>Expanding Initial Background Sample Sizes</i> .....	5-8
5.2.7	<i>Review of Background</i> .....	5-10
5.3	UPDATING BACKGROUND.....	5-12
5.3.1	<i>When to Update</i> .....	5-12
5.3.2	<i>How to Update</i> .....	5-12
5.3.3	<i>Impact of Retesting</i> .....	5-14
5.3.4	<i>Updating When Trends are Apparent</i> .....	5-14

This chapter discusses the importance and use of background data in groundwater monitoring. Guidance is provided for the proper identification, review, and periodic updating of background. Key questions to be addressed include:

- ❖ How should background be established and defined?
- ❖ When should existing background data sets be reviewed?
- ❖ How and when should background be updated?
- ❖ What impact does retesting have on background updating?

### 5.1 IMPORTANCE OF BACKGROUND

High quality background data is the single most important key to a successful statistical groundwater monitoring program, especially for detection monitoring. All of the statistical tests listed in the RCRA regulations are predicated on having *appropriate* and *representative* background measurements. As indicated in **Chapter 3**, a statistical sample is representative if the distribution of the sample measurements best follows the distribution of the population from which the sample is drawn. Representative background data has a similar but slightly different connotation. The most important quality of background is that it reflects the historical conditions unaffected by the activities it is designed to be compared to. These conditions could range from an uncontaminated aquifer to an historically contaminated site baseline unaffected by recent RCRA-actionable contaminant releases. Representative background data will therefore have numerical characteristics closely matching those arising from the site-specific aquifer being evaluated.

Background must also be *appropriate* to the statistical test. All RCRA detection monitoring tests involve comparisons of compliance point data against background. If natural groundwater conditions

have changed over time — perhaps due to cycles of drought and recharge — background measurements from five or ten years ago may not reflect current uncontaminated conditions. Similarly, recent background data obtained using improved analytical methods may not be comparable to older data. In each case, older background data may have to be discarded in favor of more recent measurements in order to construct an *appropriate* comparison. If intrawell tests are utilized due to strong evidence of spatial variability, traditional upgradient well background data will not provide an appropriate comparison. Even if the upgradient measurements are reflective of uncontaminated groundwater, appropriate background data must be obtained from each compliance point well. The main point is that compliance samples should be tested against data which best can represent background conditions now and those likely to occur in the future.

### 5.1.1 TRACKING NATURAL GROUNDWATER CONDITIONS

Background measurements, especially from upgradient wells, can provide essential information for other than formal statistical testing. For one, background data can be used to gauge mean levels and develop estimates of variability in naturally occurring groundwater constituents. They can also be used to confirm the presence or absence of anthropogenic or non-naturally occurring constituents in the site aquifer. Ongoing sampling of upgradient background wells provides a means of tracking natural groundwater conditions. Changes that occur in parallel between the compliance point and background wells may signal site-wide aquifer changes in groundwater quality not specifically attributable to onsite waste management. Such observed changes may also be indicative of analytical problems due to common artifacts of laboratory analysis (*e.g.*, re-calibration of lab equipment, errors in batch sample handling, *etc.*), as well as indications of groundwater mounding, changes in groundwater gradients and direction, migration of contaminants from other locations or offsite, etc.

Fixed GWPS like maximum contaminant levels [MCLs] may be contemplated for compliance/assessment monitoring or corrective action. Background data analysis is important if it is suspected that naturally occurring levels of the constituent(s) in question are higher than the standards or if a given hazardous constituent does not have a health- or risk-based standard. In the first case, concentrations in downgradient wells may indeed exceed the standard, but may not be attributable to onsite waste management if natural background levels *also* exceed the standard. The Parts 264 and 258 regulations recognize these possibilities, and allow for GWPS to be based on background levels.

## 5.2 ESTABLISHING AND REVIEWING BACKGROUND

Establishing appropriate background depends on the statistical approach contemplated (*e.g.*, interwell *vs.* intrawell). This section outlines the major considerations concerning how to select and develop background data including monitoring constituents and sample sizes, statistical assumptions, and the presence of data outliers, spatial variation or trends. Expanding and reviewing background data are also discussed.

### 5.2.1 SELECTING MONITORING CONSTITUENTS AND ADEQUATE SAMPLE SIZES

Due to the cost of management, mobilization, field labor, and especially laboratory analysis, groundwater monitoring can be an expensive endeavor. The most efficient way to limit costs and still meet environmental performance requirements is to minimize the total number of samples which must be sampled and analyzed. This will require tradeoffs between the number of monitoring constituents

chosen, and the frequency of background versus compliance well testing. The number of compliance wells and annual frequency of testing also affect overall costs, but are generally site-specific considerations. By limiting the number of constituents and ensuring adequate background sample sizes, it is possible to select certain statistical tests which help minimize future compliance (and total) sample requirements.

Selection of an appropriate number of detection monitoring constituents should be dictated by the knowledge of waste or waste leachate composition and the corresponding groundwater concentrations. When historical background data are available, constituent choices may be influenced by their statistical characteristics. A few representative constituents or analytes may serve to accurately assess the potential for a release. These constituents should stem from the regulated wastes, be sufficiently mobile, stable and occur at high enough concentrations to be readily detected in the groundwater. Depending on the waste composition, some non-hazardous organic or inorganic indicator analytes may serve the same purpose. The guidance suggests that between 10-15 formal detection monitoring constituents should be adequate for most site conditions. Other constituents can still be reported but not directly incorporated into formal detection monitoring, especially when large simultaneously analyzed suites like ICP-trace elements, volatile or semi-volatile organics data are run. The focus of adequate background and future compliance test sample sizes can then be limited to the selected monitoring constituents.

The RCRA regulations do not consistently specify how many observations must be collected in background. Under the Part 265 Interim Status regulations, four quarterly background measurements are required during the first year of monitoring. Recent modifications to Part 264 for Subtitle C facilities require a sequence of at least four observations to be collected in background during an interval approved by the Regional Administrator. On the other hand, at least four measurements must be collected from each background well during the first semi-annual period along with at least one additional observation during each subsequent period, for Subtitle D facilities under Part 258. Although these are minimum requirements in the regulations, are they adequate sample sizes for background definition and use?

Four observations from a population are rarely enough to adequately characterize its statistical features; statisticians generally consider sample sizes of  $n \leq 4$  to be insufficient for good statistical analysis. A decent population survey, for example, requires several hundred and often a few to several thousand participants to generate accurate results. Clinical trials of medical treatments are usually conducted on dozens to hundreds of patients. In groundwater tests, such large sample sizes are a rare luxury. However, it is feasible to obtain small sample sets of up to  $n = 20$  for individual background wells, and potentially larger sample sizes if the data characteristics allow for pooling of multiple well data.

The Unified Guidance recommends that a minimum of at least 8 to 10 independent background observations be collected before running most statistical tests. Although still a small sample size by statistical standards, these levels allow for minimally acceptable estimates of variability and evaluation of trend and goodness-of fit. However, this recommendation should be considered a temporary minimum until additional background sampling can be conducted and the background sample size enlarged (see further discussions below).

Small sample sizes in background can be particularly troublesome, especially in controlling statistical test false positive and negative rates. False negative rates in detection monitoring, *i.e.*, the

statistical error of failing to identify a real concentration increase above background, are in part a function of sample size. For a fixed false positive test rate, a smaller sample size results in a higher false negative rate. This means a decreased probability (*i.e.*, *statistical power*) that real increases above background will be detected. With certain parametric tests, control of the false positive rate using very small sample sets comes at the price of extremely low power. Power may be adequate using a non-parametric test, but control of the false positive can be lost. In both cases, increased background sample sizes result in better achievable false positive and false negative errors.

The overall recommendation of the guidance is to establish background sample sizes as large as feasible. The final tradeoff comes in the selection of the type of detection tests to be used. Prediction limit, control chart, and tolerance limit tests can utilize very small future sample sizes per compliance well (in some cases a single initial sample), but require larger background sample sizes to have sufficient power. Since background samples generally are obtained from historical data sets (plus future increments as needed), total annual sample sizes (and costs) can be somewhat minimized in the future.

### 5.2.2 BASIC ASSUMPTIONS ABOUT BACKGROUND

Any background sample should satisfy the key statistical assumptions described in **Chapter 3**. These include statistical independence of the background measurements, temporal and spatial stationarity, lack of statistical outliers, and correct distribution assumptions of the background sample when a parametric statistical approach is selected. How independence and autocorrelation impact the establishment of background is presented below, with additional discussions on outliers, spatial variability and trends in the following sections. Stationarity assumptions are considered both in the context of temporal and spatial variation.

Both the Part 264 and 258 groundwater regulations require statistically independent measurements (**Chapter 2**). Statistical *independence* is indicated by random data sets. But randomness is only demonstrated by the presence of mean and variance *stationarity* and the lack of evidence for effects such as *autocorrelation*, *trends*, *spatial and temporal variation*. These tests (described in **Part II** of this guidance) generally require at least 8 to 10 separate background measurements.

Depending on site groundwater velocity, too-frequent sampling at any given background well can result in highly *autocorrelated*, non-independent data. Current or proposed sampling frequencies can be tested for autocorrelation or other statistical dependence using the diagnostic procedures in **Chapter 14**. Practically speaking, the best way to ensure some degree of statistical independence is to allow as much time as possible to elapse between sampling events. But a balance must be drawn between collecting as many measurements as possible from a given well over a specified time period, and ensuring that the sample measurements are statistically independent. If significant dependence is identified in already collected background, the interval between sampling events may need to be lengthened to minimize further autocorrelation. With fewer sampling events per evaluation period, it is also possible that a change in statistical method may be needed, say from analysis of variance [ANOVA], which requires at least 4 new background measurements per evaluation, to prediction limits or control charts, which may require new background only periodically (*e.g.*, during a biennial update).

### 5.2.3 OUTLIERS IN BACKGROUND

Outliers or observations not derived from the same population as the rest of the sample violate the basic statistical assumption of identically-distributed measurements. The Unified Guidance recommends that testing of outliers be performed on background data, but they generally not be removed unless some basis for a likely error or discrepancy can be identified. Such possible errors or discrepancies could include data recording errors, unusual sampling and laboratory procedures or conditions, inconsistent sample turbidity, and values significantly outside the historical ranges of background data. Management of potential outliers carries both positive and negative risks, which should be carefully understood.

If an outlier value with much higher concentration than other background observations is not removed from background prior to statistical testing, it will tend to increase both the background sample mean and standard deviation. In turn, this may substantially raise the magnitude of a parametric prediction limit or control limit calculated from that sample. A subsequent compliance well test against this background limit will be much less likely to identify an exceedance. The same is true with non-parametric prediction limits, especially when the maximum background value is taken as the prediction limit. If the maximum is an outlier not representative of the background population, few truly contaminated compliance wells are likely to be identified by such a test, lowering the statistical power of the method and the overall quality of the statistical monitoring program.

Because of these concerns, it may be advisable at times to remove high-magnitude outliers in background even if the reasons for these apparently extreme observations are not known. The overall impact of removal will tend to improve the power of prediction limits and control charts, and thus result in a more environmentally protective program.

But strategies that involve automated evaluation and removal of outliers may unwittingly eliminate the evidence of real and important changes to background conditions. An example of this phenomenon may have occurred during the 1970s in some early ozone depletion measurements over Antarctica (<http://www.nas.nasa.gov/About/Education/Ozone/history.html>). Automated computer routines for outlier detection apparently removed several measurements indicating a sharp reduction in ozone concentrations, and thus prevented identification of an enlarging ozone hole by many years. Later review of the raw observations revealed that these automated routines had statistically classified measurements as outliers, which were more extreme than most of the data from that time period. Thus, there is some merit in saving and revisiting apparent 'outliers' in future investigations, even if removed from present databases.

In groundwater data collection and testing, background conditions may not be static over time. Caution should be observed in removing observations which may signal a change in natural groundwater quality. Even when conditions have not changed, an apparently extreme measurement may represent nothing more than a portion of the background distribution that has yet to be observed. This is particularly true if the background data set contains fewer than 20 samples.

In balancing these contrasting risks in retaining or removing one or more outliers, analyses of historical data patterns can sometimes provide more definitive information depending on the types of analytes and methods. For example, if a potential order-of magnitude higher outlier is identified in a sodium data set used as a monitoring constituent, cation-anion balances can help determine if this change is geochemically probable. In this case, changes to other intrawell ions or TDS should be

observed. Similarly, if a trace element outlier is identified in a single well sampling event and occurred simultaneously with other trace element maxima measured using the same analytical method (e.g., ICP-AES) either in the same well or groups of wells, an analytical error should be strongly suspected. On the other hand, an isolated increase without any other evidence could be a real but extreme background measurement. Ideally, removal of one or more statistically identified outliers should be based on other technical information or knowledge which can support that decision.

#### 5.2.4 IMPACT OF SPATIAL VARIABILITY

In the absence of contamination, comparisons made between upgradient-to-downgradient wells assume that the concentration distribution is *spatially stationary* across the well field (**Chapter 3**). This implies that every well should have the same population mean and variance, unless a release occurs to increase the concentration levels at one or more compliance wells. At many sites, this is not the case for many naturally occurring constituents. Natural or man-made differences in mean levels — referred to as *spatial variability* or *spatial variation* — impact how background must be established.

Evidence of spatial variation should drive the selection of an *intrawell* statistical approach if observed among wells known to be uncontaminated (e.g., among a group of upgradient background locations). Lack of spatial mean differences and a common variance allow for *interwell* comparisons. Appropriate background differs between the two approaches.

With interwell tests, background is derived from distinct, initially upgradient background wells, which may be enhanced by data from historical compliance wells also shown not to exhibit significant mean and variance differences. Future data from each of these compliance wells are then tested against this common background. On the other hand, intrawell background is derived from and represents historical groundwater conditions in each individual compliance well. When the population mean levels vary across a well field, there is little likelihood that the upgradient background will provide an appropriate comparison by which to judge any given compliance well.

Although spatial variability impacts the choice of background, it does so *only* for those constituents which evidence spatial differences across the well field. Each monitoring constituent should be evaluated on its own statistical merits. Spatial variation in some constituents (e.g., common ions and inorganic parameters) does not preclude the use of interwell background for other infrequently detected or non-naturally occurring analytes. At many sites, a mixture of statistical approaches may be appropriate: interwell tests for part of the monitoring list and intrawell tests for another portion. Distinct background observation sets will need to be developed under such circumstances.

Intrawell background measurements should be selected from the available historical samples at each compliance well and should include only those observations thought to be uncontaminated. Initially, this might result in very few measurements (e.g., 4 to 6). With such a small background sample, it can be very difficult to develop an adequately powerful intrawell prediction limit or control chart, even when *retesting* is employed (**Chapter 19**). Thus, additional background data will be needed to augment the testing power. One option is to periodically augment the existing background data base with recent compliance well samples (discussed in a further section below). Another possible remedy is to *statistically augment* the available sample data by running an analysis of variance [ANOVA] simultaneously on all the sets of intrawell background from the various upgradient and compliance wells (see **Chapter 13**). The *root mean squared error* [RMSE] from this procedure can be used in place of the

background standard deviation in parametric prediction and control limits to substantially increase the *effective background sample size* of such tests, despite the limited number of observations available per well.

This strategy will only work if the key assumptions of ANOVA can be satisfied (**Chapter 17**), particularly the requirement of equal variances across wells. Since natural differences in mean levels often correspond to similar differences in variability, a transformation of the data will often be necessary to homogenize the variances prior to running the ANOVA. For some constituents, no transformation may work well enough to allow the RMSE to be used as a replacement estimate for the intrawell background standard deviation. In that case, it may not be possible to construct reasonably powerful intrawell background limits until background has been updated once or twice (see **Section 5.3**).

### 5.2.5 TRENDS IN BACKGROUND

A key implication of the independent and identically distributed assumption [*i.i.d.*] is that a series of sample measurements should be *stationary over time* (*i.e.*, stable in mean level and variance). Data that are trending upward or downward violate this assumption since the mean level is changing. Seasonal fluctuations also violate this assumption since both the mean and variance will likely oscillate. The proper handling of trends in background depends on the statistical approach and the cause of the trend. With interwell tests and a common (upgradient) background, a trend can signify several possibilities:

- ❖ Contaminated background;
- ❖ A ‘break-in’ period following new well installation;
- ❖ Site-wide changes in the aquifer;
- ❖ Seasonal fluctuations, perhaps on the order of several months to a few years.

If upgradient well background becomes contaminated, intrawell testing may be needed to avoid inappropriate comparisons. Groundwater flow patterns should also be re-examined to determine if gradients are properly defined or if groundwater mounding might be occurring. With newly-installed background wells, it may be necessary to discard initially collected observations and to wait several months for aquifer disturbances due to well construction to stabilize. Site-wide changes in the underlying aquifer should be identifiable as similar trends in both upgradient and compliance wells. In this case, it might be possible to remove a common trend from both the background and compliance point wells and to perform interwell testing on the *trend residuals*. However, professional statistical assistance may be needed to do this correctly. Another option would be to switch to intrawell *trend tests* (**Chapter 17**).

Seasonal fluctuations in interwell background which are also observed in compliance wells, can be accommodated by modeling the seasonal trend and removing it from all background and compliance well data. Data seasonally-adjusted in this way (see **Chapter 14** for details) will generally be less variable than the unadjusted measurements and lead to more powerful tests than if the seasonal patterns had been ignored. For this adjustment to work properly, the same seasonal trend should be observed across the well field and not be substantially different from well to well.

Roughly linear trends in *intrawell* background usually signify the need to switch from an intrawell prediction limit or control chart to an explicit trend test, such as *linear regression* or the *Mann-Kendall* (**Chapter 17**). Otherwise the background variance will be overestimated and biased on the high side, leading to higher than expected and ultimately less powerful prediction and control limits. Seasonal fluctuations in intrawell background can be treated in one of two ways. A *seasonal Mann-Kendall* trend test built to accommodate such fluctuations can be employed (**Section 14.3.4**). Otherwise, the seasonal pattern can be estimated and removed from the background data, leaving a set of seasonally-adjusted data to be analyzed with either a prediction limit or control chart. In this latter approach, the same seasonal pattern needs to be extrapolated *beyond* the current background to more recent measurements from the compliance well being tested. These later observations also need to be seasonally-adjusted prior to comparison against the adjusted background, even if there is not enough compliance data yet collected to observe the same seasonal cycles.

When trends are apparent in background, another option is to modify the groundwater monitoring list to include only those constituents that appear to be temporally stable. Only certain analytes may indicate evidence of trends or seasonal fluctuations. More powerful statistical tests might be constructed on constituents that appear to be stationary. All such changes to the monitoring list and method of testing may require approval of the Regional Administrator or State Director.

#### 5.2.6 EXPANDING INITIAL BACKGROUND SAMPLE SIZES

In the initial development of a detection monitoring statistical program under a permit or other legal mechanism, a period of review will identify the appropriate monitoring constituents. For new sites with no prior data, plans for initial background definition need to be developed as part of permit conditions. A more typical situation occurs for interim status or older facilities which have already collected substantial historical data in site monitoring wells. For the most part, the suggestions below cover ways of expanding background data sets from existing information.

Under the RCRA interim status regulations, only a single upgradient well is required as a minimum. Generally speaking, a single background well will not generate observations that are adequately representative of the underlying aquifer. A single background well draws groundwater from only one possible background location. It is accordingly not possible to determine if spatial variation is occurring in the upgradient aquifer. In addition, a single background well can only be sampled so often since measurements that are collected too frequently run the risk of being autocorrelated. Background observations collected from a single well are typically neither representative nor constitute a large enough sample to construct powerful, accurate statistical tests. One way to expand background is to install at least 3-4 upgradient wells and collect additional data under permit.

The early RCRA regulations also allowed for aliquot replicate sampling as a means of expanding background and other well sample sizes. This approach consisted of analyzing splits or aliquots of single water quality samples. As indicated in **Chapter 2**, this approach is not recommended in the guidance. Generally limited analytical variability does not adequately capture the overall variation based on independent water quality sample data, and results in incorrect estimates of variability and degrees of freedom (a function of sample size).

Existing historical groundwater well data under consideration will need to meet the assumptions discussed earlier in this chapter— independence, stationarity, etc., including using statistical methods

which can deal with outliers, spatial and temporal variation including trends. Presuming these conditions are met, it is statistically desirable to develop as large a background sample size as practical. But no matter how many measurements are utilized, a larger sample size is advantageous only if the background samples are both appropriate to the tests selected and representative of baseline conditions.

In limited situations, upgradient-to-downgradient, interwell comparisons may be determined to be appropriate using ANOVA testing of well mean differences. To ensure appropriate and representative background, other conditions may also need to be satisfied when data from separate wells are pooled. First, each background well should be screened at the same hydrostratigraphic position as other background wells. Second, the groundwater chemistry at each of these wells should be similar. This can be checked via the use of standard geochemical bar charts, pie charts, and tri-linear diagrams of the major constituent groundwater ions and cations (Hem, 1989). Third, the *statistical* characteristics of the background wells should be similar — that is, they should be *spatially stationary*, with approximately the same means and variances. These conditions are particularly important for major water quality indicators, which generally reflect aquifer-specific characteristics. For infrequently detected analytes (e.g., filtered trace elements like chromium, silver, and zinc), even data collected from wells from *different* aquifers and/or geologic strata may be statistically indistinguishable and also eligible for pooling on an interwell basis.

If a one-way ANOVA (**Chapter 13**) on the set of background wells finds significant differences in the mean levels for some constituents, and hence, evidence of spatial variability, the guidance recommends using intrawell tests. The data gathered from the background wells will generally not be used in formal statistical testing, but are still invaluable in ensuring that appropriate background is selected.<sup>1</sup> As indicated in the discussions above and **Chapter 13**, it may be possible to pool constituent data from a number of upgradient and/or compliance wells having a common variance when parametric assumptions allow, even if mean differences exist.

When larger historical databases are available, the data can be reviewed and diagnostically tested to determine which observations best represent natural groundwater conditions suitable for future comparisons. During this review, *all historical well data* collected from both upgradient and compliance wells can be evaluated for potential inclusion into background. Wells suspected of prior contamination would need to be excluded, but otherwise each uncontaminated data point adds to the overall statistical picture of background conditions at the site and can be used to enlarge the background database. Measurements can be preferentially selected to establish background samples, so long as a consistent rationale is used (e.g., newer analytical methods, substantial outliers in a portion of a data set, etc.) Changes to an aquifer over time may require selecting newer data representing current groundwater quality over earlier results even if valid.

---

<sup>1</sup> If the spatial variation is ignored and data are pooled across wells with differing mean levels (and perhaps variances) to run an interwell parametric prediction limit or control chart test, the pooled standard deviation will tend to be substantially larger than expected. This will result in a higher critical limit for the test. Using pooled data with spatial variation will also tend to increase observed maximum values in background, leading to higher and less powerful non-parametric prediction limit tests. In either application, there will be a loss of statistical power for detecting concentration changes at individual compliance wells. Compliance wells with naturally higher mean levels will also be more frequently determined to exceed the limit than expected, while real increases at compliance wells with naturally lower means will go undetected more often.

### 5.2.7 REVIEW OF BACKGROUND

As mentioned above, if a large historical database is available, a critical review of the data can be undertaken to help establish initially appropriate and representative background samples. We recommend that other reviews of background also take place periodically. These include the following situations:

- ❖ When periodically updating background, say every 1-2 years (see **Section 5.3**)
- ❖ When performing a 5-10 year permit review

During these reviews, all observations designated as background should be evaluated to ensure that they still adequately reflect current natural or baseline groundwater conditions. In particular, the background samples should be investigated for apparent trends or outliers. Statistical outliers may need to be removed, especially if an error or discrepancy can be identified, so that subsequent compliance tests can be improved. If trends are indicated, a change in the statistical method or approach may be warranted (see earlier section on “Trends in Background”).

If background has been updated or enlarged since the last review, and is being utilized in parametric tests, the assumption of normality (or other distributional fit) should be re-checked to ensure that the augmented background data are still consistent with a parametric approach. The presence of non-detects and multiple reporting limits (especially with changes in analytical methods over time) can prove particularly troublesome in checking distribution assumptions. The methods of **Chapters 10** “Fitting Distributions” and **Chapter 15** “Handling Non-Detects” can be consulted for guidance.

Other periodic checks of the revised background should also be conducted, especially in relation to accumulated knowledge from other sites regarding analyte concentration patterns in groundwater. The following are potential sources for comparison and evaluation:

- ❖ reliable regional groundwater data studies or investigations from nearby sites;
- ❖ published literature; EPA or other agency groundwater databases like STORET;
- ❖ knowledge of typical patterns for background inorganic constituents and trace elements. An example is found in **Table 5-1** at the end of this chapter. Typical surface and groundwater levels for filtered trace elements can also be found in the published literature (*e.g.*, Hem, 1989).

Certain common features of routine groundwater monitoring analytes summarized in **Table 5-1** have been observed in Region 8 and other background data sets, which can have implications for statistical applications. Common water quality indicators like cations and anions, pH, TDS, specific conductance are almost always measurable (detectable) and generally have limited within-well variability. These would be more amenable to parametric applications; however, these measurable analytes are also most likely to exhibit well-to-well spatial variation and various kinds of within- and between-well temporal variation including seasonal and annual trends. Many of these within-well analytes are highly correlated, and would not meet the criterion for independent data if simultaneously used as monitoring constituents.

A second level of common indicator analytes— nitrate/nitrite species, fluoride, TOC and TOX—are less frequently detected and subject to more analytical detection instability (higher and lower

detection/quantitation limits). As such, these analyte data are somewhat less reliable. There is less likelihood of temporal variation, although they can exhibit spatial well differences.

Among routinely monitored .45 $\mu$ -filtered trace elements, different groups stand out. Barium is routinely detected with limited variation within most wells, but does exhibit spatial variation. Arsenic and selenium commonly occur in groundwater as oxyanions, and data can range from virtually non-detectable to always detected in different site wells. The largest group of trace elements can be considered colloidal metals (Sb, Al, Be, Cd, Cr, Co, Fe, Hg, Mn, Pb, Ni, Sn, Tl, V and Zn). While Al, Mn and Fe are more commonly detected, variability is often quite high; well-to-well spatial variability can occur at times. The remaining colloidal metals are solubility-limited in most background groundwater, generally <1 to < 10  $\mu$ g/l. But even with filtration, some natural colloidal geologic solid materials can often be detected in individual samples. Since naturally occurring Al, Mn and Fe soil solid levels are much higher, the effects on measured groundwater levels are more pronounced and variable. For most of the analytically and solubility-limited colloidal metals, there may not be any discernible well spatial differences. Often these data can be characterized by a site-wide lognormal distribution, and may be possible to pool individual well data to form larger background sizes.

With unfiltered trace element data, it is more difficult to generalize even regarding background data. The method of well sample extraction and the aquifer characteristics will determine how much solids material may be present in the samples. Excessive amounts of sample solids can result in higher levels of detection but also elevated average values and variability even for solubility-limited trace elements. The effect is most clearly seen when TSS is simultaneously collected with unfiltered data. Increases are proportional to the amount of TSS and the natural background levels for trace elements in soil/solid materials. It is recommended that TSS always be simultaneously monitored with unfiltered trace elements.

Most trace organic monitoring constituents are absent or non-detectable under clean background conditions. However, with existing up-gradient sources, it is more difficult to generalize. More soluble constituents like benzene or chlorinated hydrocarbons may be amenable to parametric distributions, but changes in groundwater levels or direction can drastically affect observed levels. For sparingly soluble compounds like polynuclear aromatics (e.g., naphthalene), aquifer effects can result in highly variable data less amenable to statistical applications.

Table 5-1 was based on the use of analytical methods common in the 1990's to the present. Detectable filtered trace element data for the most part were limited by the available analytic techniques, generally SW-846 Method 6010 ICP-AES and select AA (atomic absorption) methods with lower detection limits in the 1-10 ppb range. As newer methods are incorporated (particularly Method 6020 ICP-MS capable of parts-per-trillion detection limits for trace elements), higher quantification frequencies may result in data demonstrating more complex spatial and temporal characteristics. Table 5-1 merely provides a rough guide to where various data patterns might occur. Any extension of these patterns to other facility data sets should be determined by the formal guidance tests in **Part II**.

The background database can also be specially organized and summarized to examine common behavior among related analytes (e.g., filtered trace elements using ICP-AES) either over time or across wells during common sampling events. Parallel time series plots (**Chapter 9**) are very useful in this regard. Groups of related analytes can be graphed on the same set of axes, or groups of nearby wells for the same analyte. With either plot, highly suspect sampling events can be identified if a similar spike in

concentration or other unusual pattern occurs simultaneously at all the wells or in all the analytes. Analytical measurements that appear to be in error might be removed from the background database.

Cation-anion balances and other more sophisticated geochemical analysis programs can also be used to evaluate the reliability of existing water quality background data. A suite of tests like linear or non-parametric correlations, simple or non-parametric ANOVA described in later chapters offer overall methods for evaluating historical data for background suitability.

## 5.3 UPDATING BACKGROUND

Due both to the complex behavior of groundwater and the need for sufficiently large sample sizes, background once obtained should not be regarded as a single fixed quantity. Background should be sampled regularly throughout the life of the facility, periodically reviewed and revised as necessary. If a site uses traditional, upgradient-to-downgradient comparisons, it might seem that updating of background is conceptually simple: collect new measurements from each background well at each sampling event and add these to the overall background sample. However, significant trends or changes in one or more upgradient wells might indicate problems with individual wells, or be part of a larger site-wide groundwater change. It is worthwhile to consider the following principles for updating, whether interwell or intrawell testing is used.

### 5.3.1 WHEN TO UPDATE

There are no firm rules on how often to update background data. The Unified Guidance adopts the general principle that updating should occur when enough new measurements have been collected to allow a two-sample statistical comparison between the existing background data and a potential set of newer data. As mentioned in the following section, trend testing might also be used. With quarterly sampling, at least 4 to 8 new measurements should be gathered to enable such a test; this implies that updating would take place every 1-2 years. With semi-annual sampling, the same principle would call for updating every 2-3 years.

Updating should generally not occur more frequently, since adding a new observation to background every one or two sampling rounds does not allow a statistical evaluation of whether the background mean is stationary over time. Enough new data needs to be collected to ensure that a test of means (or medians in the case of non-normal data) can be conducted. Adding individual observations to background can introduce subtle trends that might go undetected and ultimately reduce the statistical power of formal monitoring tests.

Another practical aspect is that when background is updated, all statistical background limits (*e.g.*, prediction and control limits) needs to be recomputed to account for the revised background sample. At complex sites, updating the limits at each well and constituent on the monitoring list may require substantial effort. This includes resetting the cumulative sum [CUSUM] portions of control charts to zero after re-calculating the control limits and prior to additional testing against those limits. Too-frequent updating could thereby reduce the efficacy of control chart tests.

### 5.3.2 HOW TO UPDATE

Updating background is primarily a concern for intrawell tests, although some of the guidelines apply to interwell data. The common (generally upgradient) interwell background pool can be tested for

trends and/or changes at intervals depending on the sampling frequencies identified above. Those recently collected measurements from the background well(s) can be added to the existing pool if a Student's *t*-test or Wilcoxon rank-sum test (**Chapter 16**) finds no significant difference between the two groups at the  $\alpha = 0.01$  level of significance. Individual background wells should also be evaluated in the same manner for their respective newer data. Two-sample tests of the interwell background data are conducted to gauge whether or not background groundwater conditions have changed substantially since the last update, and are *not* tests for indicating a potential release under detection monitoring. A significant *t*-test or Wilcoxon rank-sum result should spur a closer investigation and review of the background sample, in order to determine which observations are most representative of the current groundwater conditions.

With intrawell tests using prediction limits or control charts, updating is performed both to enlarge initially small well-specific background samples and to ensure that more recent compliance measurements are not already impacted by a potential release (even if not triggered by the formal detection monitoring tests). A finding of significance using the above two-sample tests means that the most recent data *should not* be added to intrawell background. However, the same caveat as above applies: these are not formal tests for determining a potential release and the existing tests and background should continue to be used.

Updating intrawell background should also not occur until at least 4 to 8 new compliance observations have been collected. Further, a potential update is predicated on there being no *statistically significant increase* [SSI] recorded for that well constituent, including since the last update. Then a *t*-test or Wilcoxon rank-sum comparison can be conducted at each compliance well between existing intrawell background and the potential set of newer background. A non-significant result implies that the newer compliance data can be re-classified as background measurements and added to the existing intrawell background sample. On the other hand, a determination of significance suggests that the compliance observations should be reviewed to determine whether a gradual trend or other change has occurred that was missed by the intervening prediction limit or control chart tests. If intrawell tests make use of a common pooled variance, the assumption of equal variance in the pooled wells should also be checked with the newer data.

Some users may wish to evaluate historical and future background data for potential trends. If plots of data versus time suggest either an overall trend in the combined data sets or distinct differences in the respective sets, linear or non-parametric trend tests covered in **Chapter 17** might be used. A determination of a significant trend might occur even if the two-sample tests are inconclusive, but individual group sample sizes should be large enough to avoid identifying a significant trend based on too few samples and perhaps randomly occurring. A trend in the newer data may reflect or depart from the historical data conditions. Some form of statistical adjustments may be necessary, but see **Section 5.3.4** below.

### 5.3.3 IMPACT OF RETESTING

A key question when updating intrawell background is how to handle the results of retesting.<sup>2</sup> If a retest confirms an SSI, background should not be updated. Rather, some regulatory action at the site should be taken. But what if an initial exceedance of a prediction or control limit is *disconfirmed* by retesting? According to the logic of retesting (**Chapter 19**), the well passes the compliance test for that evaluation and monitoring should continue as usual. But what should be done with the initial exceedance when it comes time to update background at the well?

The initial exceedance may be due to a laboratory error or other anomaly that has caused the observation to be an outlier. If so, the error should be documented and not included in the updated background sample. But if the exceedance is not explainable as an outlier or error, it may represent a portion of the background population that has heretofore not been sampled. In that case, the data value could be included in the updated background sample (along with the repeat sample) as evidence of the expanded but true range of background variation. Ultimately, it is important to characterize the background conditions at the site as completely and accurately as possible, so as to minimize both false positive and false negative decision errors in compliance testing.

The severity and classification of the initial exceedance will depend on the specific retesting strategy that has been implemented (**Chapter 19**). Using the same background data in a parametric prediction limit or control chart test, background limits are proportionately lower as the 1-of- $m$  order increases (higher  $m$ ). Thus, a 1-of-4 prediction limit will be lower than a 1-of-3 limit, and similarly the 1-of-3 limit lower than for a 1-of-2 test. An initial exceedance triggered by a 1-of-4 test limit and disconfirmed by a repeat sample, might not trigger a lower order prediction limit test. The initial sample value may represent an upper tail value from the true distribution. Retesting schemes derive much of their statistical power by allowing more frequent initial exceedances, even if some of these represent possible measurements from background. The initial and subsequent resamples *taken together* are designed to identify which initial exceedances truly represent SSIs and which do not. These tests presume that occasional excursions beyond the background limit will occur. Unless the exceedance can be documented as an outlier or other anomaly, it should probably be included in the updated intrawell background sample.

### 5.3.4 UPDATING WHEN TRENDS ARE APPARENT

An increasing or decreasing trend may be apparent between the existing background and the newer set of candidate background values, either using a time series plot or applying **Chapter 17** trend analyses. Should such trend data be added to the existing background sample? Most detection monitoring tests assume that background is stationary over time, with no discernible trends or seasonal variation. A mild trend will probably make very little difference, especially if a Student- $t$  or Wilcoxon rank-sum test between the existing and candidate background data sets is non-significant. More severe or continuing trends are likely to be flagged as SSIs by formal intrawell prediction limit or control chart tests.

---

<sup>2</sup> With interwell tests, the common (upgradient) background is rarely affected by retests at compliance point wells (unless the latter were included in the common pool). Should retesting fail to confirm an initial exceedance, the initial value can be reported alongside the disconfirming resamples in statistical reports for that facility.

With interwell tests, a stronger trend in the common upgradient background may signify a change in natural groundwater quality across the aquifer or an incomplete characterization of the full range of background variation. If a change is evident, it may be necessary to delete some of the earlier background values from the updated background sample, so as to ensure that compliance testing is based on current groundwater conditions and not on outdated measures of groundwater quality.

**Table 5-1. Typical Background Data Patterns for Routine Groundwater Monitoring Analytes**

Analyte Groups	Detection Rates		Between Well Mean Differences	Within Well Variability (CVs)	Between Well Equal Variances	Outlier Problems	Temporal Variation					Typical Distribution within well	Data Grouping
	Frequency of Detection by Well	Multiple Reporting Limits					Between Well by Analyte	Within Well	Within Well	Within Well	Within Well		
<b>Inorganic Constituents and Indicators</b>													
Major ions, pH, TDS, Specific Conductance	High to 100%		✓✓✓	Generally low (.1-.5)	✓✓	✓	✓✓	✓✓✓	✓✓	✓✓	✓✓	Normal	Intrawell
CO3, F, NO2,NO3	Some to most detectable	✓✓	✓✓	Moderate (.2-1.5)	Variable	✓✓	✓			✓	✓	Norm, Log or NPM	Intrawell/ Interwell
<b>.45µ Filtered Trace Elements</b>													
Ba	High to 100%	✓✓	✓✓✓	Low (.1-.5)	✓	✓	✓				✓	Normal	Intrawell
As, Se	Some wells high, others low to zero	✓✓	✓✓ (some wells)	Moderate (.2-1.5)	Variable	✓✓	✓				✓	Normal, Log or NPM	Intrawell/ Interwell
Al, Mn, Fe	Low to Moderate	✓✓	✓	Moderate to high (.3->2.0)	✓	✓✓✓	✓				✓	Log or NPM	Intrawell/ Interwell
Sb, Be, Cd, Cr, Cu, Hg, Pb, Ni, Ag, Tl, V, Zn	Zero to low	✓✓✓		Moderate to high (.5->2.0)	✓✓	✓✓✓	✓	✓✓			✓	Log or NPM	Interwell or NDC
<b>Trace Organic and Indicator Analytes (patterns at sites with prior contamination; generally absent in clean sites)</b>													
VOA's-BETX and Cl-Hydrocarbons	Variable, can be high	✓	Variable by site and wells			✓	Variable by site and specific wells					Normal, Log or NPM	Intrawell, Interwell or NDC
BNAs, Other Trace Organics	Generally low-mod	✓✓	" " " " "			✓	" " " "					" "	" "
Indicators: TOX, TPH, TOC, sulfide	Variable	✓✓	" " " " "			✓✓✓	" " " "					" "	" "

NPM- non-parametric methods; NDC- never-detected constituents

**Checks:** None- unknown, absent or infrequently occurring; ✓ - Occasionally; ✓✓ - Frequently; ✓✓✓ - Very Frequently