Designation: D7048 – 16

Standard Guide for Applying Statistical Methods for Assessment and Corrective Action Environmental Monitoring Programs[1]

This standard is issued under the fixed designation D7048; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 The scope and purpose of this guidance is to present a variety of statistical approaches for assessment, compliance and corrective action environmental monitoring programs. Although the methods provided here are appropriate and often optimal for many environmental monitoring problems, they do not preclude use of other statistical approaches that may be equally or even more useful for certain site-specific applications.

1.2 In the following sections, the details of select statistical procedures used in assessment and corrective action programs for environmental monitoring (soil, groundwater, air, surface water, and waste streams) are presented.

1.3 The statistical methodology described in the following sections should be used as guidance. Other methods may also be appropriate based on site-specific conditions or for monitoring situations or media that are not presented in this document.

1.4 This practice offers an organized collection of information or a series of options and does not recommend a specific course of action. This document cannot replace education, experience and professional judgements. Not all aspects of this practice may be applicable in all circumstances. This ASTM standard is not intended to represent or replace the standard of care by which the adequacy of a given professional service must be judged without consideration of a project's many unique aspects. The word Standard in the title of this document only means that the document has been approved through the ASTM consensus process.

1.5 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety and health practices and determine the applicability of regulatory requirements prior to use.*

2. Referenced Documents

2.1 *ASTM Standards:*[2]

D653 Terminology Relating to Soil, Rock, and Contained Fluids

D5092 Practice for Design and Installation of Groundwater Monitoring Wells

D5792 Practice for Generation of Environmental Data Related to Waste Management Activities: Development of Data Quality Objectives

D6250 Practice for Derivation of Decision Point and Confidence Limit for Statistical Testing of Mean Concentration in Waste Management Decisions

D6312 Guide for Developing Appropriate Statistical Approaches for Groundwater Detection Monitoring Programs

3. Terminology

3.1 *Definitions*—For definitions of common terms in this guid, see Terminology D653.

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *corrective action monitoring*—under RCRA (in the United States), corrective action monitoring is instituted when hazardous constituents from a RCRA regulated unit have been detected at statistically significant concentrations between the compliance point and the downgradient facility property boundary as specified under 40 CFR 264.100. Corrective action monitoring is conducted throughout a corrective action program that is implemented to address groundwater contamination. At non-RCRA sites, corrective action monitoring is conducted throughout the active period of corrective action to determine the progress of remediation and to identify statistically significant trends in groundwater contaminant concentrations.

3.2.2 *false positive rate*—the rate at which the statistical procedure indicates contamination when contamination is not present.

[1] This guide is under the jurisdiction of ASTM Committee D18 on Soil and Rock and is the direct responsibility of Subcommittee D18.21 on Groundwater and Vadose Zone Investigations.

Current edition approved Oct. 1, 2016. Published October 2016. Originally approved in 2004. Last previous edition approved in 2010 as D7048–04(2010). DOI: 10.1520/D7048-16.

[2] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

3.2.3 *lognormal distribution*—a frequency distribution whose logarithm follows a normal distribution.

3.2.4 *lower confidence limit, LCL*—a lower limit that has a specified probability (for example, 95 %) of including the true concentration (or other parameter). Taken together with the upper confidence limit, forms a confidence interval that will include the true concentration with confidence level that accounts for both tail areas (for example, 90 %).

3.2.5 *lower prediction limit, LPL*—a statistical estimate of the minimum concentration that will provide a lower bound for the next series of k measurements from that distribution, or the mean of m new measurements for each of k sampling locations, with specified level of confidence (for example, 95 %).

3.2.6 *nonparametric*—a term referring to a statistical technique in which the distribution of the constituent in the population is unknown and is not restricted to be of a specified form.

3.2.7 *nonparametric prediction limit*—the largest (or second largest) of n background samples. The confidence level associated with the nonparametric prediction limit is a function of n, m and k.

3.2.8 *normal distribution*—a frequency distribution whose plot is a continuous, infinite, bell-shaped curve that is symmetrical about its arithmetic mean, mode and median (which are numerically equivalent). The normal distribution has two parameters, the mean and variance.

3.2.9 *outlier*—a measurement that is statistically inconsistent with the distribution of other measurements from which it was drawn.

3.2.10 *parametric*—a term referring to a statistical technique in which the distribution of the constituent in the population is assumed to be known.

3.2.11 *potential area of concern*—areas with a documented release or likely presence of a hazardous substance that could pose an unacceptable risk to human health or the environment.

3.2.12 *upper confidence limit, UCL*—an upper limit that has a specified probability (for example, 95 %) of including the true concentration (or other parameter). Taken together with the lower confidence limit, the UCL forms a confidence interval that will include the true concentration with confidence level that accounts for both tail areas.

3.2.13 *upper prediction limit, UPL*—a statistical estimate of the maximum concentration that will not be exceeded by the next series of k measurements from that distribution, or the mean of m new measurements for each of k sampling locations, with specified level of confidence (for example, 95 %) based on a sample of n background measurements.

3.3 *Symbols:* μ = the true population mean of a constituent

$\bar{x}$ = the sample-based mean or average concentration of a constituent computed from n background measurements which differs from μ because of sampling variability, and other error

σ^2 = the true population variance of a constituent

s^2 = the sample-based variance of a constituent computed from n background measurements

s = the sample-based standard deviation of a constituent computed from n background measurements

$\bar{y}$ = the mean of the natural log transformed data (also the natural log of the geometric mean)

s_y = the standard deviation of the natural log transformed data

n = the number of background (offsite or upgradient) measurements

k = the number of future comparisons for a single monitoring event (for example, the number of downgradient monitoring wells multiplied by the number of constituents to be monitored) for which statistics are to be computed

α = the false positive rate for an individual comparison (that is, one sampling location and constituent)

m = the number of onsite or downgradient measurements used in computing the onsite mean concentration

α^* = the site-wide false positive rate covering the sampling locations and constituents

t = the $100(1 - \alpha)$ percentage point of Student's t-distribution on $n - 1$ degrees of freedom

H_L = the factor developed by Land (1971) (**1**)[3] to obtain the lower $100(\alpha)$ % confidence limit for the mean of a lognormal distribution

H_U = the factor developed by Land (1971) (**1**) to obtain the upper $100(\alpha)$ % confidence limit for the mean of a lognormal distribution

4. Summary of Guide

4.1 The guide is summarized as Figs. 1-7. These figures provides a flow-chart illustrating the steps used in computing the comparisons to regulatory or health based groundwater protection standard (GWPS) in assessment and corrective action environmental monitoring programs.

5. Significance and Use

5.1 The principal use of this standard is in assessment, compliance and corrective action environmental monitoring programs (for example, for a facility that could potentially contaminate groundwater). The significance of the guidance is that it presents a statistical method that allows comparison of groundwater data to regulatory and/or health based limits.

5.2 Of course, there is considerable support for statistical methods applied to detection, assessment and corrective action monitoring programs that can be applied to environmental sites.

NOTE 1—For example, in the United States, the 90 % upper confidence limit (UCL) of the mean is used in USEPA's SW846 (Chapter 9) for determining if a waste is hazardous. If the UCL is less than the criterion for a particular hazardous waste code, then the waste is not a hazardous waste even if certain individual measurements exceed the criterion. Similarly, in the USEPA Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities Addendum to the Interim Final Guidance (1992) (**2**), confidence intervals for the mean and various upper percentiles of the distribution are advocated for assessment and corrective action. Interestingly, both the 1989 and 1992 USEPA guidance documents (**2, 3**) suggest use of the lower 95 % confidence limit (LCL) as a tool for

[3] The boldface numbers in parentheses refer to a list of references at the end of this standard.

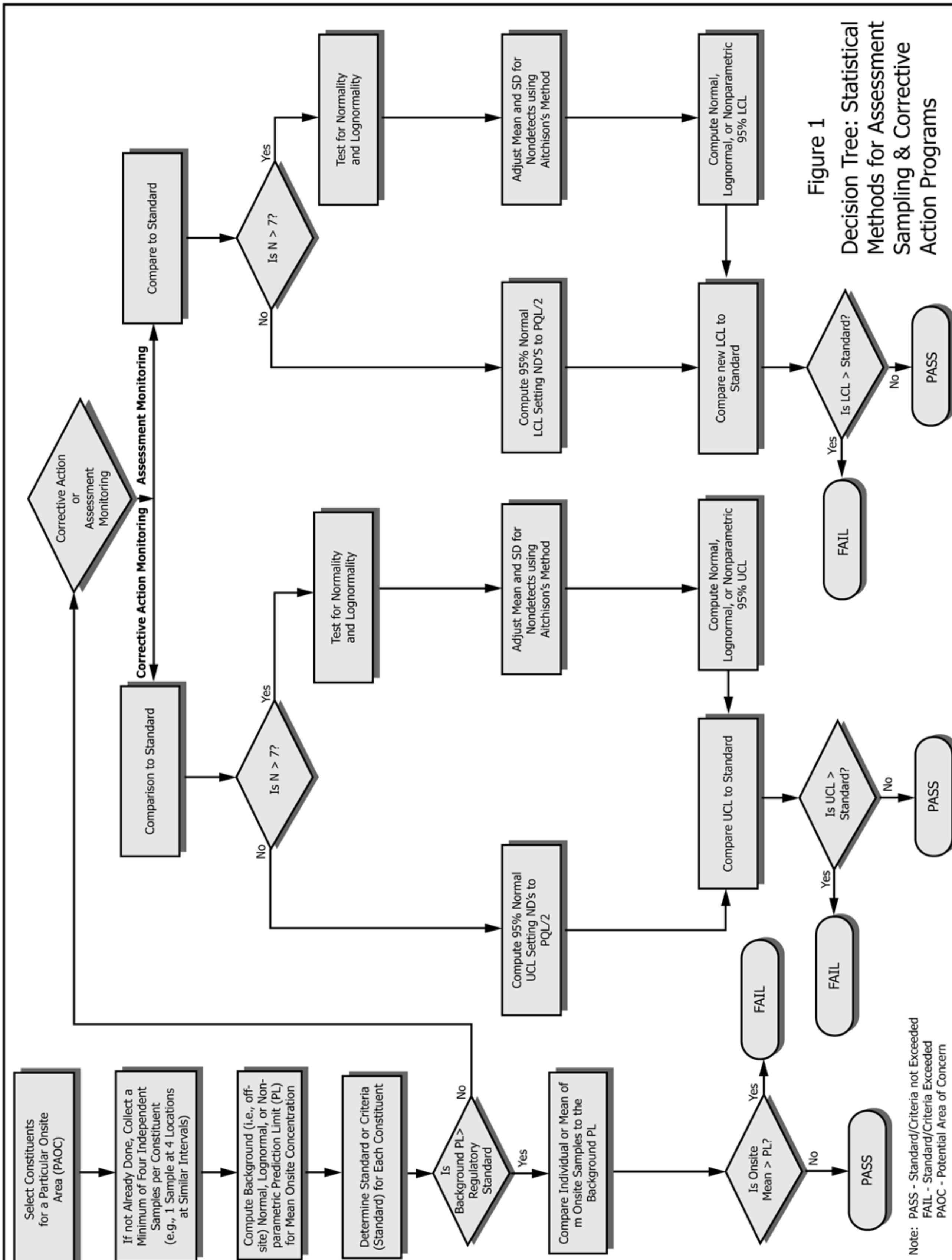

FIG. 1 Decision Tree—Statistical Methods for Assessment Sampling and Corrective Action Programs

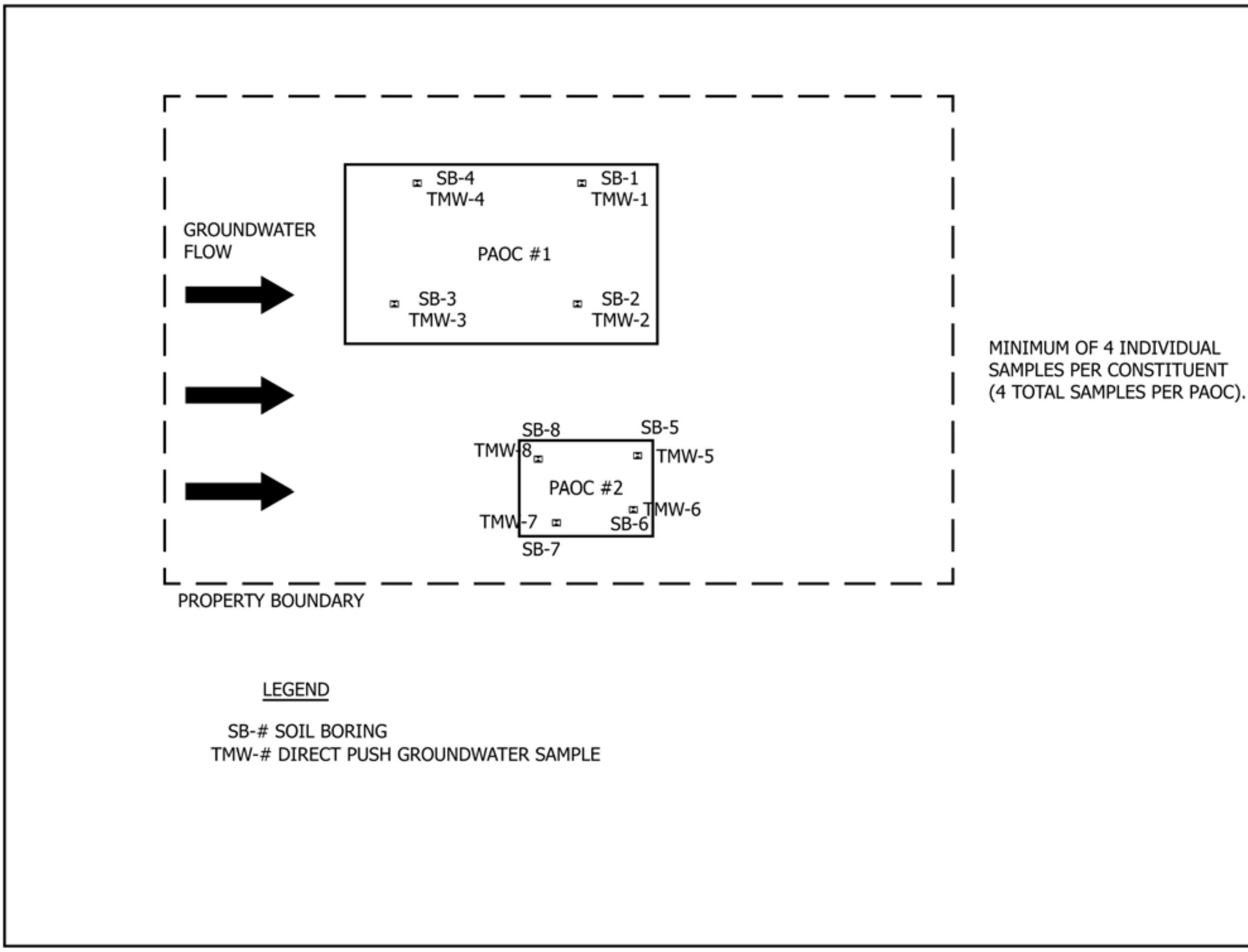

FIG. 2 Single PAOC Comparison to a Standard/Criteria

determining whether a criterion has been exceeded in assessment monitoring.

The latest guidance in this area calls for use of the LCL in assessment monitoring and the UCL in corrective action. In this way, corrective action is only triggered if there is a high degree of confidence that the true concentration has exceeded the criterion or standard, whereas corrective action continues until there is a high degree of confidence that the true concentration is below the criterion or standard. This is the general approach adopted in this guide, as well.

5.3 There are several reasons why statistical methods are needed in assessment and corrective action monitoring programs. First, a single measurement indicates very little about the true concentration in the sampling location of interest, and with only one sample it cannot be determined if the measured concentration is a typical or an extreme value. The objective is to compare the true concentration (or some interval that contains it) to the relevant criterion or standard. Second, in many cases the constituents of interest are naturally occurring (for example, metals) and the naturally existing concentrations may exceed the relevant criteria. In this case, the relevant comparison is to background (for example, off-site soil or upgradient groundwater) and not to a fixed criterion. As such, background data should be statistically characterized to obtain a statistical estimate of an upper bound for the naturally occurring concentrations so that it can be confidently determined if onsite concentrations are above background levels. Third, there is often a need to compare numerous potential constituents of concern to criteria or background, at numerous sampling locations. By chance alone there will be exceedances as the number of comparisons becomes large. The statistical approach to this problem can decrease the potential for false positive results.

5.4 Statistical methods for detection monitoring have been well studied in recent years (see Gibbons, 1994a, 1996, USEPA 1992 (**2, 4, 5**) and Practice D6312, formerly PS 64-96 authored by Gibbons, Brown and Cameron, 1996). Although equally important, statistical methods for assessment monitoring, Phase I and II Investigations, on-going monitoring and corrective action monitoring have received less attention, (Gibbons and Coleman, 2001) (**6**).

5.5 The guide is summarized in Fig. 1, which provides a flow-chart illustrating the steps in developing a statistical evaluation method for assessment and corrective action programs. Fig. 1 illustrates the various decision points at which the

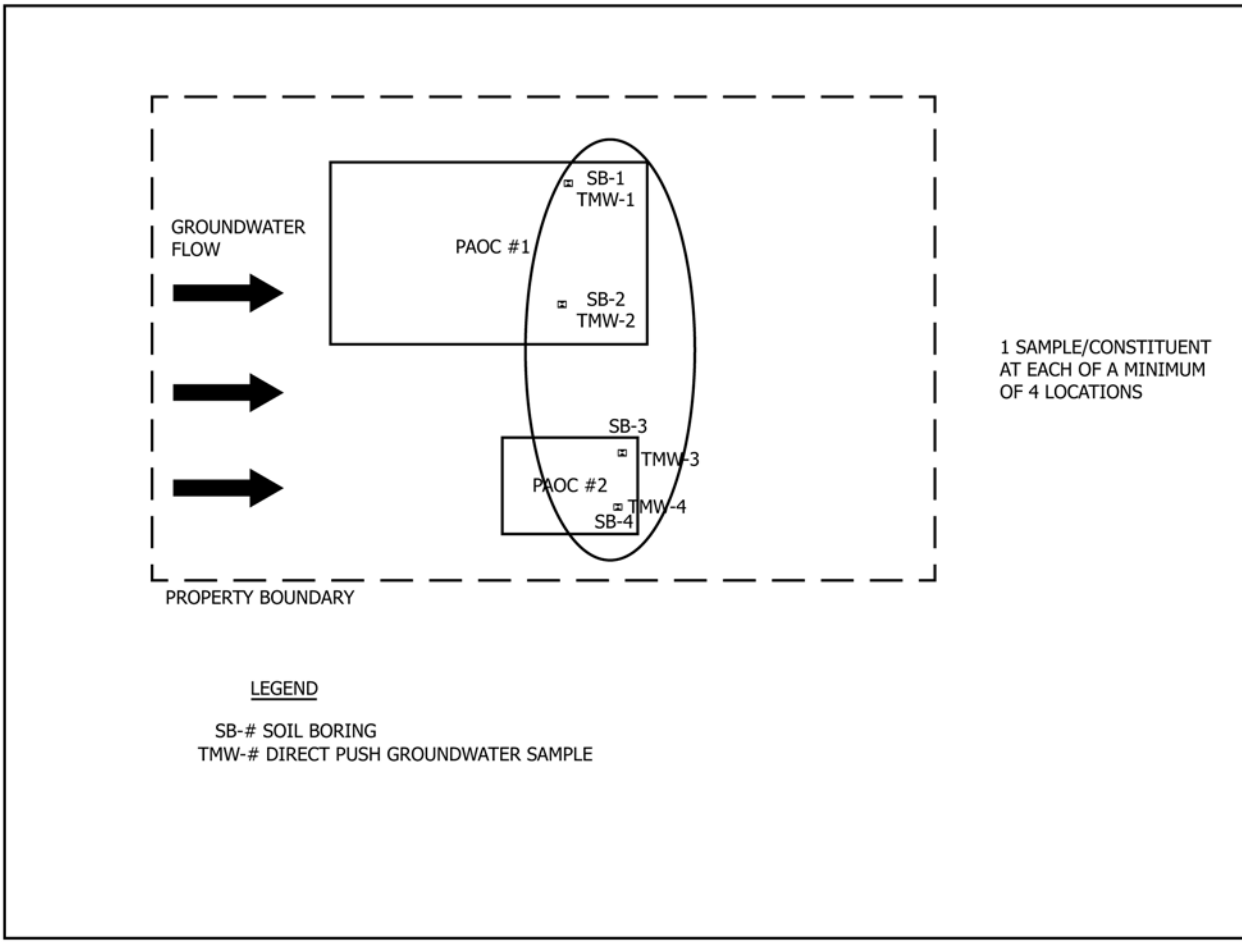

FIG. 3 Multiple PAOC Comparison to a Standard/Criteria

general comparative strategy is selected, and how the statistical methods are to be selected based on site-specific considerations.

6. Procedure

6.1 In the following, the general conceptual and statistical foundations of the sampling program are described. Following this general discussion, media-specific details (that is, soil, groundwater, and waste streams) are provided.

6.1.1 Identify relevant constituents for the specific type of facility, media (for example, soil and/or groundwater) and area of interest. A facility is generally comprised of a series of subunits or "source areas" that may have a distinct set of sampling locations and relevant constituents of concern (referred to as a PAOC). The subunit may consist of a single sampling point or collection of sampling points. In some cases, the entire site may comprise the area of interest and all sampling locations are considered jointly. The boundaries of the "source area" or "decision unit" should be defined. In most cases, the owner/operator should select the smallest practical list of constituents that adequately characterize the source area in terms of historical use.

6.1.2 For each constituent obtain the appropriate regulatory criterion or standard (for example, maximum contaminant level, MCL) if one is available. The appropriate criterion or standard should be selected based on relevant pathways (for example, direct contact, ingestion, inhalation) and appropriate land use criteria (for example, commercial, industrial, residential).

6.1.3 For each constituent which may have a background concentration higher than the relevant health based criterion, set "background" to the upper 95 % confidence prediction limit (UPL) as described in the Technical Details section. The prediction limits are computed from available data collected from background, or outside source areas that are unlikely to be contaminated, upstream, upwind or upgradient locations only. Henceforth, background refers to these types of offsite sources. The background data are first screened for outliers and then tested for normality and lognormality (see Technical Details section).

6.1.3.1 If the test of normality cannot be rejected (for example, at the 95 % confidence level), background is equal to the 95 % confidence normal prediction limit.

6.1.3.2 If the test of normality is rejected but the test of lognormality cannot be rejected, background is equal to the 95 % confidence lognormal prediction limit.

6.1.3.3 If the data are neither normal nor lognormal, or the detection frequency is less than 50 %, background is the nonparametric prediction limit. When we are interested in a single potentially impacted measurement, normal, lognormal,

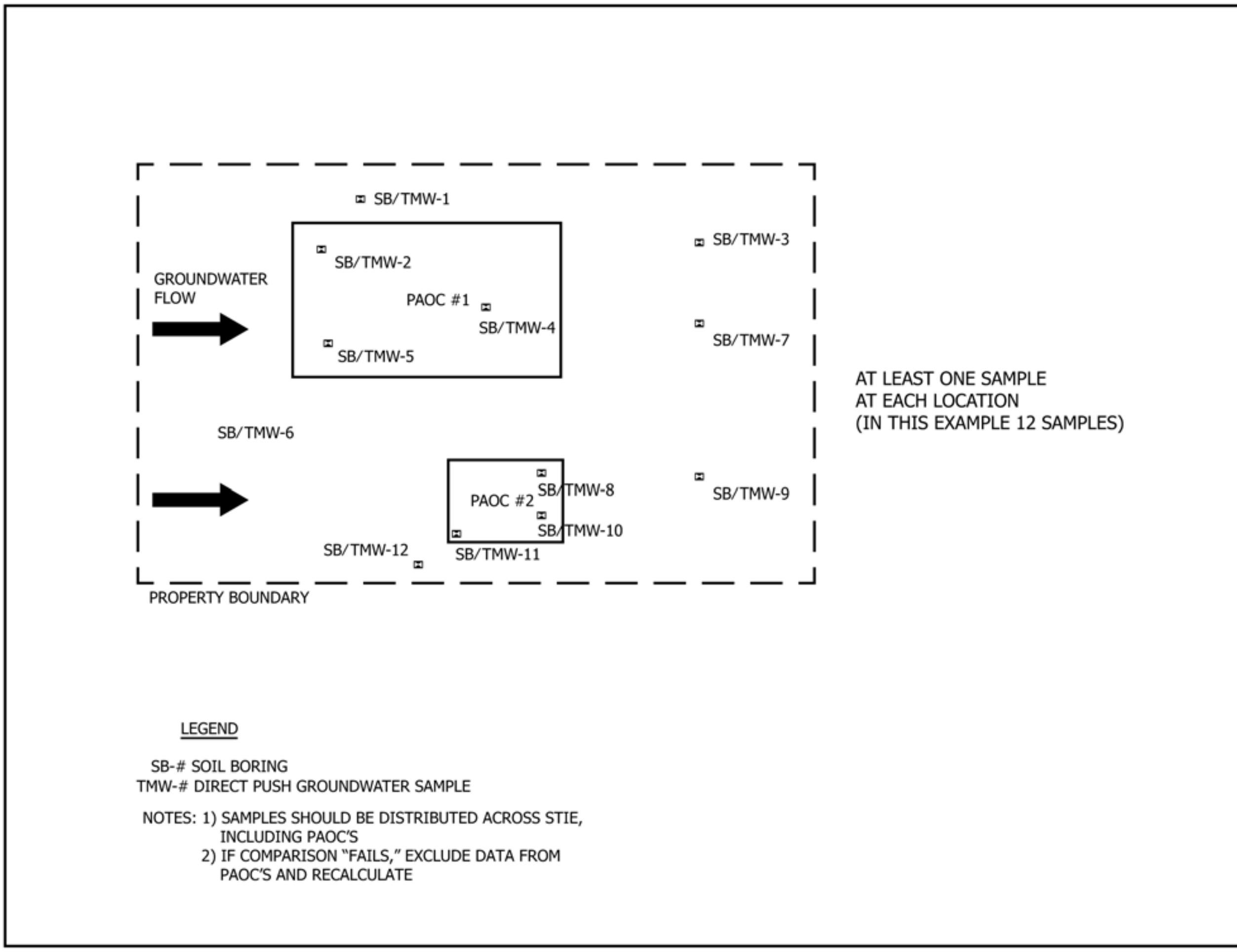

FIG. 4 Comparison of Mean Concentrations of Entire Site to a Standard/Criteria

and nonparametric prediction limits are identical with respect to the parameter being compared (that is, an individual measurement). However, when the comparison to background is for an onsite/downgradient mean concentration, they differ in that the nonparametric prediction limit is for the median whereas the parametric prediction limits are for the mean. This limitation is unavoidable, so whenever practical, parametric prediction limits should be used. Note that, if the detection frequency is zero, background is set equal to the appropriate Quantification Limit (QL) for that constituent which is the lowest concentration that can be reliably determined within specified limits of precision and accuracy by the indicated methods under routine laboratory operating conditions.

6.1.3.4 If the background is greater than the relevant criterion or standard or if there is no criterion or standard, then comparisons are made to the background prediction limit. If the criterion is greater than background, then compare the appropriate confidence limit to the criterion. Note that if nothing is detected in background, then the background is the QL. If the criterion is lower than the QL, then the criterion is the QL.

6.1.4 The number of samples taken depends on whether comparison is to background or a criterion and whether comparisons are made at individual locations or by pooling samples within a source area. If comparison is to background, collect one or more samples from each source area or sampling location. If comparison is to a criterion (that is, the criterion is greater than background), and interest is in a single location, four or more independent samples from each sampling location will be needed. If the comparison is to a criterion for an entire source area, one or more samples from each of four sampling locations within the source area are needed. If there are fewer than four sampling locations within a given source area, then the total number of measurements from the source area must be four or more (for example, two sampling locations each with two independent samples). Note that these sample sizes represent absolute minimum necessary for the statistical computations. In general, a larger number of samples will be needed to obtain a representative sample of the population of interest.

6.1.5 If comparison is to a criterion or standard there are two general approaches. In assessment, monitoring where interest is in determining if a criterion has been exceeded, compare the 95 % lower confidence limit (LCL) for the mean of four or more samples from a single location, source area or the entire site to the relevant criterion. In corrective action sampling and monitoring, where interest is in demonstrating that the onsite concentration is lower than the criterion, compare the 95 %

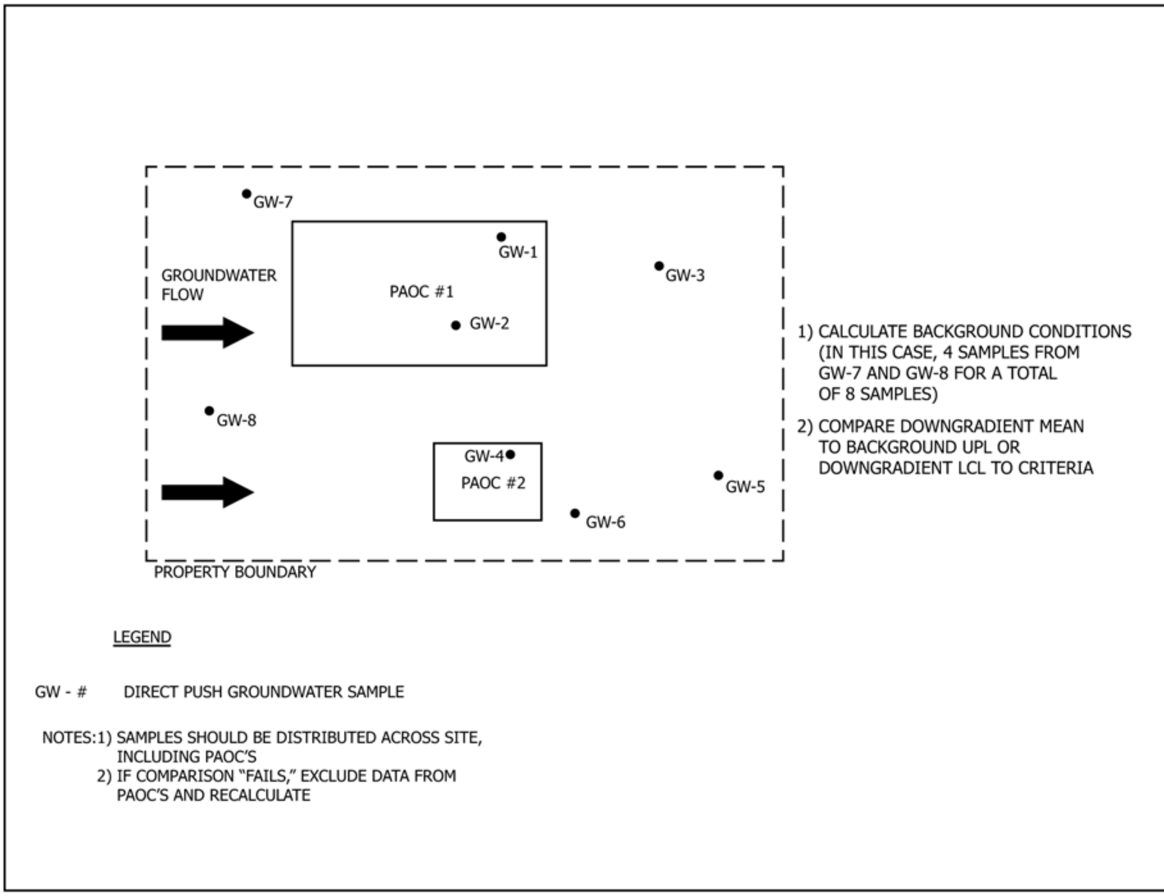

FIG. 5 Evaluation of Groundwater Concentrations for the Entire Site

upper confidence limit (UCL) for the mean of four or more samples from a single location, source area or the entire site to the relevant criterion.

6.1.6 If the background prediction limit is larger than the relevant criterion, then do one of the following: *(1)* for a single measurement obtained from an individual location, compare this individual measurement to the background prediction limit for the next single measurement from each of *k* locations, *(2)* for multiple measurements obtained from a given source area or the entire site, compare the mean of the measurements to the background prediction limit for the mean of *m* measurements based on the best fitting statistical distribution or nonparametric alternative.

6.1.7 Note that if the background UPL and the regulatory criterion are quite similar, the downgradient mean may exceed the background UPL but the LCL for the downgradient mean may still be less than the regulatory criterion. In this case, an exceedance is not determined. Fig. 1 presents a decision tree that can be used to step through the statistical analysis approach.

6.1.8 In the following sections, application to specific media and types of sampling and monitoring programs is described. The areas covered include soil, groundwater and waste stream sampling; however, similar approaches can be taken for air and surface water monitoring.

6.2 *Soils—Evaluation of Individual Source Areas (PAOCs):*

6.2.1 Collect soil samples from the surface to the groundwater table at appropriate intervals in the most likely contaminated location in the source area and screen soils to determine the interval with highest concentration(s).

6.2.2 At three or more other nearby borings located in the same source area, collect one sample in the same vertical interval (geologic profile) as the previously identified highest concentration interval (that is, the first, boring in the interval of highest screening concentration).

6.2.3 Send the samples from the vertical interval in the four borings to the lab for analysis. As in 6.1.5 these intervals and sample sizes represent a minimum needed for the statistical computations and larger numbers will typically be needed in practice to provide adequate characterization of the area of interest.

6.2.4 Compute the 95 % LCL (assessment) or UCL (corrective action) for the mean of the *m* results to determine if the particular PAOC exceeds the regulatory criterion.

6.2.5 If an exceedance is found, assess whether it is naturally occurring (for example, metals) by obtaining eight or more independent background samples (that is, offsite soil samples from the same interval) and compute the 95 % confidence upper prediction limit (UPL) for the mean of the *m* onsite/downgradient samples, and compare the UPL to the observed mean at each PAOC. An exceedance is determined only if the PAOC mean concentration exceeds both the

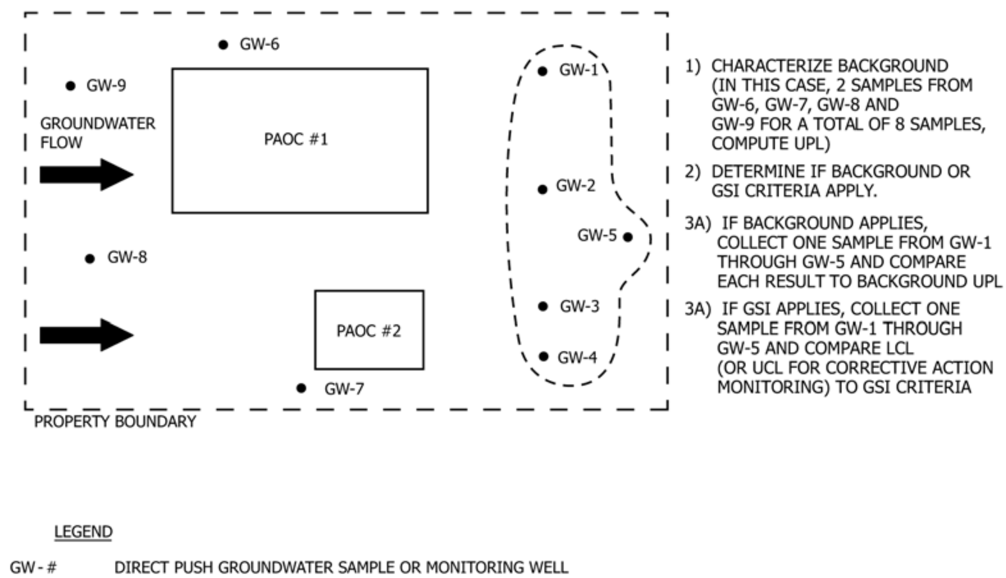

FIG. 6 Evaluation of Groundwater Data to Determine Compliance with GSI Criteria

regulatory criterion and the background UPL. Figs. 2 and 3 illustrate the sampling location approaches for this scenario. Eight samples are needed because for fewer, uncertainty in the background mean and variance will lead to unacceptably large UPLs.

6.3 *Soils—Area-Wide or Site-Wide Evaluations:*

6.3.1 Collect soil samples to be representative of the entire spatial distribution of constituents of concern (four or more samples).

6.3.2 Compute the 95 % LCL (assessment) or UCL (corrective action) for the mean of the onsite samples and determine if the area or site as a whole exceeds the regulatory criterion.

6.3.3 If an exceedance is found, check that it is not naturally occurring by obtaining eight or more independent background samples (that is, offsite soil samples from the same stratigraphic unit) and compute the 95 % confidence UPL.

6.3.4 If the level of hazardous substance concentrations at the site is relatively homogeneous, compute the UPL for the mean of the *m* onsite measurements and compare the observed mean to the UPL.

6.3.5 If the level of hazardous substance concentrations at the site is heterogeneous, compute the UPL for the *m* , individual onsite measurements and compare each measurement to the UPL.

6.3.6 An exceedance is determined only if the area or site-wide mean concentration exceeds both the regulatory criterion and the background UPL.

6.3.7 If an exceedance is found, it may be practical to exclude PAOCs one at a time until the Site minus the selected PAOCs does not exceed criterion. This method may be appropriate only when sufficient sampling of the PAOC has been conducted as part of the site or area-wide evaluation. Fig. 4 illustrates the sampling location approach for this scenario.

6.4 *Groundwater—Aquifer:*

6.4.1 As in the soil sampling above, if soil sampling and screening or prior groundwater monitoring indicates that groundwater may be impacted, then one groundwater sample will be obtained in each of four or more borings using a direct push methodology or from existing groundwater monitoring wells and results will be evaluated statistically to determine if the entire PAOC requires additional assessment. The general methodology previously described for Soil PAOCs can be used here as well, as illustrated in Figs. 2 and 3. To characterize background, eight or more independent samples should be collected. This can be four samples from each of two locations, two samples from four locations, or one sample from each of eight locations. Two or more locations are needed. Statistical independence implies that the same groundwater is not

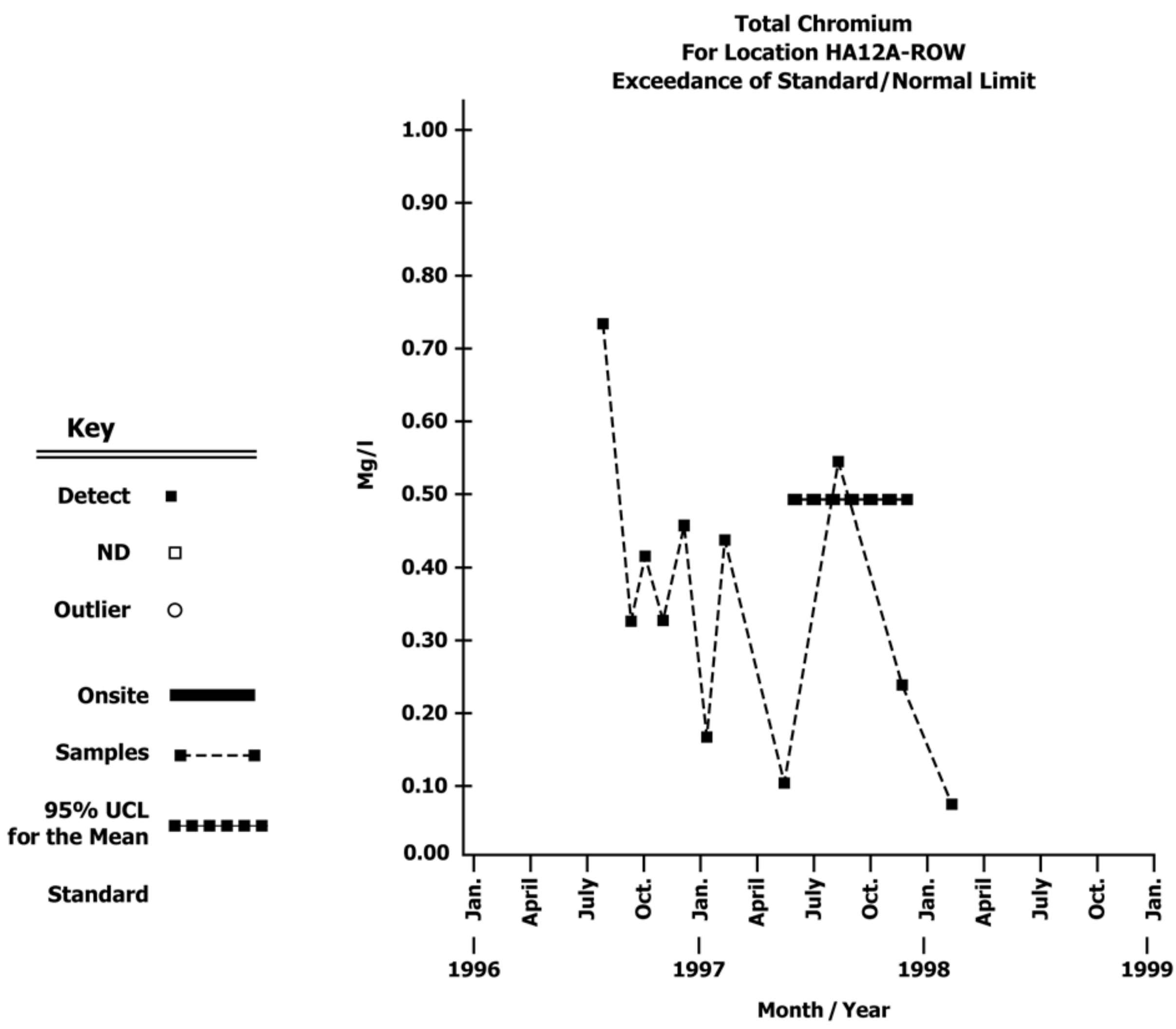

FIG. 7 Graph 1—Comparison to a Standard

sampled repeatedly and that the background data are representative of the same temporal variation as are the onsite data. This precludes establishing background in the winter and comparing onsite measurements obtained in the summer. As in previous sections, these sampling requirements represents minimum requirements of the statistical procedures and the actual numbers of samples and time frames should be based on geologic criteria (for example, see "Guide for Developing Conceptual Models for Contaminated Sites").

6.4.2 Fig. 5 illustrates another approach for evaluating groundwater at a site. Sampling locations are set up as shown and four independent samples are collected from background locations GW-7 and GW-8 (that is, eight total background samples). If the background UPL exceeds the appropriate regulatory criterion, then the mean from downgradient samples GW-1 through GW-6 is compared to the background UPL to determine if an exceedance exists. If the background UPL is less than the appropriate regulatory criterion, then the downgradient LCL should be compared to the criterion. Another modification to this approach is if an exceedance exists, the m downgradient samples can be compared individually to background to determine if the impact is restricted to a subset of monitoring locations. As previously discussed, if the background UPL and the regulatory criterion are quite similar, the downgradient mean could exceed the background UPL but the LCL for the downgradient mean may still be less than the regulatory criterion. In this case, an exceedance is not determined. This applies equally to all media.

6.5 *Groundwater—Groundwater Surface Water Interface (GSI):*

6.5.1 Characterize background. As previously indicated, background is established by obtaining eight or more independent samples from two or more locations (that is, to incorporate spatial variability). The background limit is established by computing the UPL from these data (that is, eight or more background samples).

6.5.2 If the only comparison is to background, obtain a single sample from each GSI sampling location (that is, one sample from each compliance point) and compare to the appropriate upgradient UPL.

6.5.3 If comparison is to regulatory criteria, obtain four or more independent samples from each GSI sampling location (that is, four samples from each compliance point) and compare the LCL (assessment) or the UCL (corrective action) to the regulatory criterion. If the upgradient UPL is greater than the regulatory criterion for a particular constituent, compare each GSI sampling location to background.

6.5.4 Depending on the application, each GSI sampling location can be compared to background or the appropriate regulatory criterion individually or as a group. Fig. 6 illustrates the sampling strategy for this scenario.

6.6 *Groundwater—Long-term Monitoring:*

6.6.1 When sampling for long-term monitoring of a plume, compute the 95 % normal UCL for the most recent four measurements in each sampling location and compare to the relevant regulatory criteria (see 7.2.1.4).

6.6.2 Compute Sen's test (**7**) to determine if there are increasing or decreasing trends (at a 95 % confidence level) at each sampling location (needing 8 or more measurements per well).

6.7 *Groundwater—Natural Attenuation Evaluation:*

6.7.1 Here temporal changes are considered in the mean of the wells within a plume or wells in the relatively higher concentration area of a plume.

6.7.2 Obtain eight or more independent samples (for example, one from each of eight monitoring wells or two from each of four monitoring wells). This should be done either for the wells within the plume or the relatively higher concentration area of the plume. Note that if there is seasonal variability in analyte concentrations, four quarterly samples within a period of one year or more should be obtained from each sampling location.

6.7.3 Compute the 95 % confidence lower prediction limit (LPL) and the UPL for the mean of the wells in the plume or wells within the relatively higher concentration area. For example, if there are 8 wells, compute the LPL and UPL for the mean of the next 8 samples.

6.7.4 If the actual mean exceeds the UPL, there is evidence that the plume is getting significantly worse.

6.7.5 If the actual mean is less than the LPL, there is evidence that the plume is getting significantly better (that is, natural attenuation is occurring).

6.7.6 Compute Sen's test to determine if there are increasing or decreasing trends (at a 95 % confidence level) at each sampling location (needing 8 or more measurements per well).

6.8 *Waste Stream Sampling:*

6.8.1 To determine if a particular waste stream is hazardous, obtain a series of $n \geq 4$ representative samples from the waste stream for relevant characteristically hazardous criteria.

6.8.2 Compute the appropriate 90 % UCL for the mean concentration.

6.8.3 Note that the 90 % confidence level is used based on guidance provided in SW846 Chapter 9 (US EPA).

6.8.4 If the 90 % UCL is less than the regulatory criterion or standard, the waste stream is not hazardous.

7. Technical Approach

7.1 The purpose of this section is to provide a description of the specific statistical methods to be used in assessment and corrective action sampling programs (see Gibbons and Coleman, 2001) (**6**).

7.2 *Comparison to a Regulatory Criterion or Standard:*

7.2.1 *Confidence Limits for the Mean or Median Concentration:*

7.2.1.1 The 95 % normal LCL (assessment sampling and monitoring) or 95 % normal UCL (corrective action) for the mean of four or more measurements are computed and compared to the Regulatory Criterion or Standard.

7.2.1.2 The 95 % normal LCL, for one-tailed (assessment sampling and monitoring) for the mean of m measurements is computed as:

$$\bar{x} - t_{[m-1,0.95]} \frac{s}{\sqrt{m}} \quad (1)$$

7.2.1.3 The 95 % normal UCL (corrective action) for the mean of m measurements is computed as:

$$\bar{x}+t_{[m-1,0.95]}\frac{s}{\sqrt{m}} \quad (2)$$

7.2.1.4 If $m < 8$, nondetects should be replaced either by the reported measured concentration (if available) or one-half of the QL since with fewer than eight measurements, more sophisticated statistical adjustments are not appropriate. Note that direct comparison between measured concentrations below the QL and a regulatory standard should not be made. Similarly, a normal UCL is used because seven or fewer samples are insufficient to confidently determine distributional form of the data. Use of a lognormal limit with small samples can result in extreme limit estimates, therefore default to normality for $m < 8$.

7.2.1.5 If $m \geq 8$, use Aitchison's (1955) (**8**) method to adjust for nondetects and test for normality and lognormality of the data using the single group or multiple group version of the Shapiro-Wilk test (see 7.3.2.2 for details). The multiple group version of the Shapiro-Wilk test is used when there are multiple measurements from multiple onsite locations (use 95 % confidence level). Note that alternatives such as Cohen's (1961) method can be used, however the reporting limit should be constant for each constituent, which is rarely the case.

7.2.1.6 If $m \geq 8$, and the data are neither normally nor lognormally distributed, compute the 95 % nonparametric LCL or UCL for the median of m, samples (see Hahn and Meeker (1991) section 5.2 (**9**), and Gibbons and Coleman (2001) (**6**)). Alternatively, if the data are lognormally distributed, compute a lognormal LCL or UCL for the mean (see Land, 1971) (**1**) . The (1 − α) 100 % lognormal UCL for the mean is:

$$\exp\left(\bar{y}+0.5s_y+\frac{H_{1-\alpha}s_y}{\sqrt{m-1}}\right) \quad (3)$$

The (1 − α) 100 % lognormal LCL for the mean is:

$$\exp\left(\bar{y}+0.5s_y+\frac{H_{\alpha}s_y}{\sqrt{m-1}}\right) \quad (4)$$

In general, the LCL or UCL for the mean should be used except in the nonparametric case where it is not defined. In addition, caution should be taken using Land's method in that it is not robust to departures from lognormality.

7.2.1.7 The factors H are given by Land (1975) (**10**) and $\bar{y}$ and s_y are the mean and standard deviation of the natural log transformed data (that is, $y = \log_e(\bar{x})$). The lognormal LCL or UCL for the median is simply the exponentiated result of computing the normal LCL or UCL on natural log transformed data (see Hahn and Meeker, 1991 (**9**) and Gibbons and Coleman, 2000 (**6**)).

7.2.1.8 Use Sen's nonparametric trend test to evaluate trends (either increasing or decreasing) to demonstrate the effectiveness of corrective action (see Gibbons, 1994, pp. 175-178) (**4**). The Mann Kendall test is also a valid alternative, (see Gibbons 1994 which also discusses methods for seasonal adjustments).

7.2.2 *Confidence Limits for Other Percentiles of the Distribution:*

7.2.2.1 For some applications, there may be interest in a LCL or UCL for a specific percentile of the distribution (for example, 90th, 95th or 99th percentiles of the concentration distribution). Of course, in the nonparametric case, only confidence limits for percentiles are available, such as the 50th percentile of the distribution (that is, the concept of confidence limits for the mean does not exist without a specific parametric form of the distribution).

7.2.2.2 For those constituents with short-term exposure risks or in those cases in which one may wish to show added environmental protection, confidence limits for upper percentiles of the distribution may be used (for example, 90th, 95th or 99th percentiles). The interpretation here is that there is 95 % confidence that 95 % of the distribution is beneath the estimated confidence limit. Both LCLs and UCLs for upper percentiles can be computed and normal, lognormal and nonparametric approaches have been described in general by Hahn and Meeker (1991) (**9**) and more recently by Gibbons and Coleman (2001) (**6**), and are closely related to statistical tolerance limits.

7.3 *Generation of the Background Prediction Limit:*

7.3.1 When the background upper prediction limit exceeds the Regulatory Criterion or Standard, then onsite measurements are compared to the 95 % confidence upper prediction limit based on the available background data for that constituent. In the following section, the method by which the prediction limit is computed is presented.

7.3.2 *Case 1—Compounds Quantified in All Background Samples:*

7.3.2.1 For groundwater, obtain four or more measurements from two or more background sampling locations. For soils, obtain measurements from eight or more different background sampling locations.

7.3.2.2 For groundwater, in which measurements are taken repeatedly from the same sampling location (that is, an upgradient sampling well), test normality of distribution using the multiple group version of the Shapiro-Wilk test (Wilk and Shapiro, 1968) (**11**) applied to n upgradient or background measurements. The n background measurements refer to available background measurements obtained at multiple background sampling locations (spatial) and available sampling events (temporal). The multiple group version of the original Shapiro-Wilk test (Shapiro and Wilk, 1965) (**12**) takes into consideration that upgradient measurements are nested within different upgradient sampling wells, hence the original Shapiro-Wilk test does not apply. This computation is described by Gibbons, 1994 (pp. 228-231) (**4**). For soils, the n background samples can be tested for normality using the original Shapiro-Wilk test (see Gibbons, 1994, pp. 219-222) (**4**) since each measurement is obtained from a unique background sampling location.

7.3.2.3 If normality is not rejected (that is, at the 95 % confidence level), compute the 95 % (that is, site-wide) prediction limit as:

$$\bar{x}+t_{[n-1,1-\alpha]}s\sqrt{\frac{1}{m}+\frac{1}{n}} \quad (5)$$

where:

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

$$s = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}}$$

α = false positive rate for each individual test,
$t_{[n-1, 1-\alpha]}$ = one-sided $(1 - \alpha)$ 100 % point of Student's t distribution on $n - 1$ degrees of freedom, and
n = number of background measurements, and m is the number of measurements from which the onsite source area mean is computed. Note that if individual onsite measurements are to be compared to background (for example, the most recent measurement from each location), $m = 1$ and the prediction limit becomes:

$$\bar{x} + t_{[n-1,1-\alpha]} s \sqrt{1 + \frac{1}{n}} \tag{6}$$

7.3.2.4 Select $\alpha = 0.05/k$, where k is the number of comparisons (that is, sampling locations or source areas times the number of constituents).

7.3.2.5 If normality is rejected, take natural logarithms of the n background measurements and recompute the multiple group Shapiro-Wilk test.

7.3.2.6 If the transformation results in a nonsignificant G statistic (that is, the values $\log_e(x)$ are normally distributed), compute the lognormal prediction limit as:

$$\exp\left(\bar{y} + t_{[n-1,1-\alpha]} s_y \sqrt{\frac{1}{m} + \frac{1}{n}}\right) \tag{7}$$

where:

$$\bar{y} = \sum_{i=1}^{n} \frac{\log_e(x_i)}{n} \tag{8}$$

and

$$s_y = \sqrt{\sum_{i=1}^{n} \frac{(\log_e(x_i) - \bar{y})^2}{n-1}} \tag{9}$$

7.3.2.7 For $m > 1$ this lognormal prediction limit is for the onsite geometric mean or median concentration. To compute an approximate lognormal prediction limit for the onsite arithmetic mean concentration ($m > 1$) use Land's method and compute (see Gibbons and Coleman, 2001) (**6**). Also see Bhuamik and Gibbons, 2004, (**13**) for a new and more statistically rigorous alternative.

$$\exp\left(\bar{y} + 0.5 s_y + H_{1-\alpha} s_y \sqrt{\frac{1}{m} + \frac{1}{n}}\right) \tag{10}$$

7.3.2.8 If log transformation does not bring about normality (that is, the probability of G is less than 0.01), compute the non-parametric prediction limit which is an order statistic (that is, an ordered measurement such as the maximum) of the background concentration measurements. For the case of $m = 1$, tables are provided by Gibbons (1994a, chapter 2) (**4**) for confidence levels based on using the largest ($x_{(n)}$) or second largest ($x_{(n-1)}$) measurement as the prediction limit as a function of n and k with and without verification resampling. For $m > 1$, one-sided nonparametric prediction limits for the median of m onsite measurements are given by Hahn and Meeker (1991, section 5.5.2) (**9**).

7.3.2.9 In the context of groundwater monitoring, this general decision tree is described in Practice D6312 (formerly PS64-96).

7.3.3 *Case 2—Compounds Quantified in 50 % or More of Background Samples:*

7.3.3.1 Apply the multiple group Shapiro-Wilk test to the quantified measurements only.

7.3.3.2 If the data are normally distributed compute the mean of the n background samples as:

$$\bar{x} = \left(1 - \frac{n_0}{n}\right)\bar{x}' \tag{11}$$

Where n_0 is the number of samples in which the compound was not detected, n is the total number of measurements and x' is the average of the $n - n_0$ detected values. The standard deviation is:

$$s = \sqrt{\left(1 - \frac{n_0}{n}\right)s^{2'} + \frac{n_0}{n}\left(1 - \frac{n_0 - 1}{n - 1}\right)\bar{x}^{2'}} \tag{12}$$

where s' is the standard deviation of the $n - n_0$ detected measurements. The normal prediction limit can then be computed as previously described. This method is due to Aitchison (1955) (**8**) .

7.3.3.3 If the multiple group Shapiro-Wilk test reveals that the data are lognormally distributed, replace x' with y' and s' with s'_y in the equations for x and s.

7.3.3.4 The lognormal prediction limit for the onsite mean or median concentration may then be computed as previously described.

7.3.3.5 Note that this adjustment only applies to positive random variables. The natural logarithm of concentrations less than 1 are negative and therefore the adjustment does not apply. For this reason, add 1 to each value (that is, $\log_e(x_i + 1) \geq 0$) and subtract 1 from the exponentiated limit. This generally happens because the data have been presented in units of mg/L. Converting to ppb will often eliminate this problem.

7.3.3.6 If the data are neither normally or lognormally distributed, compute a nonparametric prediction limit.

7.3.4 *Case 3—Compounds Quantified in Less than 50 % of Background Samples:*

7.3.4.1 For individual comparisons of the most recent measurement in each sampling location to background (that is, $m = 1$), the nonparametric prediction limit for the next single measurement in each of k sampling locations is the largest concentration found in n background measurements.

7.3.4.2 Gibbons (1990, 1991, 1994) (**4, 14, 15**) has shown that the confidence associated with this decision rule, is a function of the multivariate extension of the hypergeometric distribution.

7.3.4.3 Perform tabulation of confidence levels for $n = 4,..., 100$, $k = 1,..., 100$ comparisons (that is, sampling locations), is presented in Gibbons, 1994 (Table 2.5) (**4**).

7.3.4.4 To compare the source area median to background (that is, $m > 1$), compute a nonparametric prediction limit for the 50th percentile of the distribution of m onsite samples based on *n* background samples using the method described by Hahn and Meeker (1991, section 5.5) (**9**) and Gibbons and Coleman (2001) (**6**).

7.3.5 *Detection of Outliers*—From time to time anomalous results may be found among background samples due to a laboratory, sampling or clerical error. The net result is that the background prediction limit will be dramatically larger than it should be, leading to a less environmentally conservative, sampling program. To eliminate these problems, background data are screened for outliers using Dixon's test at the 99 % confidence level (see Gibbons, 1994, pp. 254-257) (**4**) .

8. Example

8.1 This example illustrates the use of statistical procedures at a site undergoing long-term monitoring. At the X Site (Site) manufacturing operations have ceased, the production equipment has been relocated, and the Site has been sold to a third party.

8.2 Hexavalent chromium was detected in soil and groundwater in the vicinity of a former chromium plating area located in the southeast quadrant of the facility. Site evaluations defined the distribution of chemical constituents and physical characteristics of this area, and an interim groundwater remediation system was designed and installed as the primary remedial action component for groundwater. The system includes a blasted bedrock trench, a groundwater recovery system with electrical controls, a remote telemetry unit (RTU), and an ion exchange groundwater treatment system which discharges treated groundwater to the sanitary sewer system.

8.3 A baseline sampling event occurred August 1 through 5, 1996, prior to remedial system start-up and permanent operation. After initiation of permanent operation, selected wells were monitored bi-weekly for the first month of operation, and quarterly thereafter. Other selected wells are monitored on an annual basis. The groundwater protection standard for total chromium is 0.05 mg/L (ppm).

8.4 Following the guidance (see 6.6), compare the 95 % UCL for the mean concentration for the last four available measurements for each monitoring well and constituent to the relevant criterion and test for increasing or decreasing trends in the data for each well and constituent using Sen's test. To illustrate the method, data for total chromium in monitoring well MW—1 were analyzed. Fig. 7 presents the graphical results of the analysis and presents step by step computational details. Review of Fig. 7 reveal that: *(a)* there is a decreasing trend in the well, which demonstrates the beneficial effects of the remediation; *(b)* the most current measurement is now close to the cleanup criterion of 0.050 mg/L; *(c)* the UCL for the mean concentration is 0.488 mg/L which is still an order of magnitude above the criterion indicating that remediation should continue.

9. Keywords

9.1 assessment monitoring; confidence limit; corrective action; environmental monitoring; prediction limit; quantification limit; statistics

REFERENCES

(1) Land, C. E., "Confidence Intervals for Linear Functions of the Normal Mean and Variance," *Annals of Mathematical Statistics*, 42, 1971, pp. 1187-1205.

(2) USEPA, Addendum to Interim, *Final Guidance Document Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities*, July, 1992.

(3) USEPA, *Statistical Analysis of Groundwater Monitoring Data at RCRA Facilities*, Unified Guidance, EPA/530/R-09-007, March 2009.

(4) Gibbons, R. D., *Statistical Methods for Groundwater Monitoring*, 2nd Edition, John Wiley and Sons, New York, October 2009.

(5) Gibbons, R. D.,"Some Conceptual and Statistical Issues in Analysis of Groundwater Monitoring Data," *Environmetrics*, Vol 7, Issue 2, March 1996, pp. 185-199.

(6) Gibbons R. D., and Coleman D. E., *Statistical Methods for Detection and Quantification of Environmental Contamination*, John Wiley and Sons, 2001.

(7) Sen, P. K., "Estimates of the Regression Coefficient Based on Kendall's Tau," *Journal of the American Statistical Association*, 63, 1968, pp. 1319-1389.

(8) Aitchison, J., "On the Distribution of a Positive, Random Variable Having a Discrete Probability Mass at the Origin," *Journal of the American Statistical Association*, 50, 1995, pp. 901-908.

(9) Hahn, G. J., and Meeker, W. Q., *Statistical Intervals: A Guide for Practitioners*, John Wiley and Sons, New York, 1991.

(10) Land, C. E., "Tables of Confidence Limits for Linear Functions of the Normal Mean and Variance," *Selected Tables of Mathematical Statistics*, Vol III, American Mathematical Society, Providence R.I., 1975, Vol 3, pp. 385-419.

(11) Wilk, M. B., and Shapiro, S. S., "The Joint Assessment of Normality of Several Independent Samples," *Technometrics*, Vol 10, 1968, pp. 825-839.

(12) Shapiro, S. S., and Wilk, M. B., "An Analysis of Variance Test for Normality(Complete Samples)," *Biometrika*, Vol 52, 1965, pp. 591-611.

(13) Bhaumik DK, Gibbons RD, An Upper Prediction Limit for the Arithmetic mean of a Lognormal Random Variable *Technometrics*, 46, 2004, pp. 239-248.

(14) Gibbons, R. D.,"A General Statistical Procedure for Groundwater Detection Monitoring at Waste Disposal Facilities," *Ground Water*, Vol 28, 1990, pp. 235-243.

(15) Gibbons, R. D.,"Some Additional Nonparametric Prediction Limits for Groundwater Monitoring at Waste Disposal Facilities," *Ground Water*, Vol 29, 1991, pp. 729-736.

(16) Gibbons, R. D.,"Statistical Models for the Analysis of Volatile Organic Compounds in Waste Disposal Facilities," *Ground Water*, Vol 25, 1987, pp. 572-580; Vol 21, 1987, pp. 292-319.