



Standard Guide for Developing Appropriate Statistical Approaches for Ground- Water Detection Monitoring Programs¹

This standard is issued under the fixed designation D 6312; 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 This guide covers the context of ground-water monitoring at waste disposal facilities, regulations have required statistical methods as the basis for investigating potential environmental impact due to waste disposal facility operation. Owner/operators must perform a statistical analysis on a quarterly or semiannual basis. A statistical test is performed on each of many constituents (for example, 10 to 50 or more) for each of many wells (5 to 100 or more). The result is potentially hundreds, and in some cases, a thousand or more statistical comparisons performed on each monitoring event. Even if the false positive rate for a single test is small (for example, 1 %), the possibility of failing at least one test on any monitoring event is virtually guaranteed. This assumes you have done the correct statistic in the first place.

1.2 This guide is intended to assist regulators and industry in developing statistically powerful ground-water monitoring programs for waste disposal facilities. The purpose of these methods is to detect a potential ground-water impact from the facility at the earliest possible time while simultaneously minimizing the probability of falsely concluding that the facility has impacted ground water when it has not.

1.3 When applied inappropriately existing regulation and guidance on statistical approaches to ground-water monitoring often suffer from a lack of statistical clarity and often implement methods that will either fail to detect contamination when it is present (a false negative result) or conclude that the facility has impacted ground water when it has not (a false positive). Historical approaches to this problem have often sacrificed one type of error to maintain control over the other. For example, some regulatory approaches err on the side of conservatism, keeping false negative rates near zero while false positive rates approach 100 %.

1.4 The purpose of this guide is to illustrate a statistical ground-water monitoring strategy that minimizes both false negative and false positive rates without sacrificing one for the other.

1.5 This guide is applicable to statistical aspects of ground-

water detection monitoring for hazardous and municipal solid waste disposal facilities.

1.6 It is of critical importance to realize that on the basis of a statistical analysis alone, it can never be concluded that a waste disposal facility has impacted ground water. A statistically significant exceedance over background levels indicates that the new measurement in a particular monitoring well for a particular constituent is inconsistent with chance expectations based on the available sample of background measurements.

1.7 Similarly, statistical methods can never overcome limitations of a groundwater monitoring network that might arise due to poor site characterization, well installation and location, sampling, or analysis.

1.8 It is noted that when justified, intra-well comparisons are generally preferable to their inter-well counterparts because they completely eliminate the spatial component of variability. Due to the absence of spatial variability, the uncertainty in measured concentrations is decreased making intra-well comparisons more sensitive to real releases (that is, false negatives) and false positive results due to spatial variability are completely eliminated.

1.9 Finally, it should be noted that the statistical methods described here are not the only valid methods for analysis of ground-water monitoring data. They are, however, currently the most useful from the perspective of balancing site-wide false positive and false negative rates at nominal levels. A more complete review of this topic and the associated literature is presented by Gibbons (1).²

1.10 The values stated in both inch-pound and SI units are to be regarded as the standard. The values given in parentheses are for information only.

1.11 *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 limitations prior to use.*

1.12 *This guide 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 or experience and should be used in conjunction with professional*

¹ This guide is under the jurisdiction of ASTM Committee D-18 on Soil and Rock and is the direct responsibility of Subcommittee D18.21 on Ground Water and Vadose Zone Investigations.

Current edition approved Sept. 10, 1998. Published December 1998.

² The boldface numbers given in parentheses refer to a list of references at the end of the text.

judgment. Not all aspects of this guide 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, nor should this document be applied without consideration of a project's many unique aspects. The word "Standard" in the title of this document means only that the document has been approved through the ASTM consensus process.

2. Terminology

2.1 Definitions:

2.1.1 *assessment monitoring program, n*—ground-water monitoring that is intended to determine the nature and extent of a potential site impact following a verified statistically significant exceedance of the detection monitoring program.

2.1.2 *combined Shewart (CUSUM) control chart, n*—a statistical method for intra-well comparisons that is sensitive to both immediate and gradual releases.

2.1.3 *detection limit (DL), n*—the true concentration at which there is a specified level of confidence (for example, 99 % confidence) that the analyte is present in the sample (2).

2.1.4 *detection monitoring program, n*—ground-water monitoring that is intended to detect a potential impact from a facility by testing for statistically significant changes in geochemistry in a downgradient monitoring well relative to background levels.

2.1.5 *intra-well comparisons, n*—a comparison of one or more new monitoring measurements to statistics computed from a sample of historical measurements from that same well.

2.1.6 *inter-well comparisons, n*—a comparison of a new monitoring measurement to statistics computed from a sample of background measurements (for example, upgradient versus downgradient comparisons).

2.1.7 *prediction interval or limit, n*—a statistical estimate of the minimum or maximum concentration, or both, that will contain the next series of k measurements with a specified level of confidence (for example, 99 % confidence) based on a sample of n background measurements.

2.1.8 *quantification limit (QL), n*—the concentration at which quantitative determinations of an analyte's concentration in the sample can be reliably made during routine laboratory operating conditions (3).

2.2 Definitions of Terms Specific to This Standard:

2.2.1 *false negative rate, n*—in detection monitoring, the rate at which the statistical procedure does not indicate possible contamination when contamination is present.

2.2.2 *false positive rate, n*—in detection monitoring, the rate at which the statistical procedure indicates possible contamination when none is present.

2.2.3 *nonparametric, adj*—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.

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

2.2.5 *parametric, adj*—a term referring to a statistical technique in which the distribution of the constituent in the

population is assumed to be known.

2.2.6 *verification resample, n*—in the event of an initial statistical exceedance, one (or more) new independent sample is collected and analyzed for that well and constituent which exceeded the original limit.

2.3 Symbols:

2.3.1 α —the false positive rate for an individual comparison (that is, one well and constituent).

2.3.2 α^* —the site-wide false positive rate covering all wells and constituents.

2.3.3 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.

2.3.4 n —the number of background measurements.

2.3.5 σ^2 —the true population variance of a constituent.

2.3.6 s —the sample-based standard deviation of a constituent computed from n background measurements.

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

2.3.8 μ —the true population mean of a constituent.

2.3.9 \bar{x} —the sample-based mean or average concentration of a constituent computed from n background measurements.

3. Summary of Guide

3.1 This guide is summarized in Figs. 1, that provides a flowchart illustrating the steps in developing a statistical monitoring plan. The monitoring plan is based either on background versus monitoring well comparisons (for example, upgradient versus downgradient comparisons or intra-well comparisons, or a combination of both). Fig. 1 illustrates the various decision points at which the general comparative strategy is selected (that is, upgradient background versus intra-well background) and how the statistical methods are to be selected based on site-specific considerations. The statistical methods include parametric and nonparametric prediction limits for background versus monitoring well comparisons and combined Shewart-CUSUM control charts for intra-well comparisons. Note that the background database is intended to expand as new data become available during the course of monitoring.

4. Significance and Use

4.1 The principal use of this guide is in ground-water detection monitoring of hazardous and municipal solid waste disposal facilities. There is considerable variability in the way in which existing Guide USEPA regulation and guidance are interpreted and practiced. Often, much of current practice leads to statistical decision rules that lead to excessive false positive or false negative rates, or both. The significance of this proposed guide is that it jointly minimizes false positive and false negative rates at nominal levels without sacrificing one error for another (while maintaining acceptable statistical power to detect actual impacts to ground-water quality (4)).

4.2 Using this guide, an owner/operator or regulatory agency should be able to develop a statistical detection monitoring program that will not falsely detect contamination when it is absent and will not fail to detect contamination when it is present.

Development of a Statistical Detection Monitoring Plan

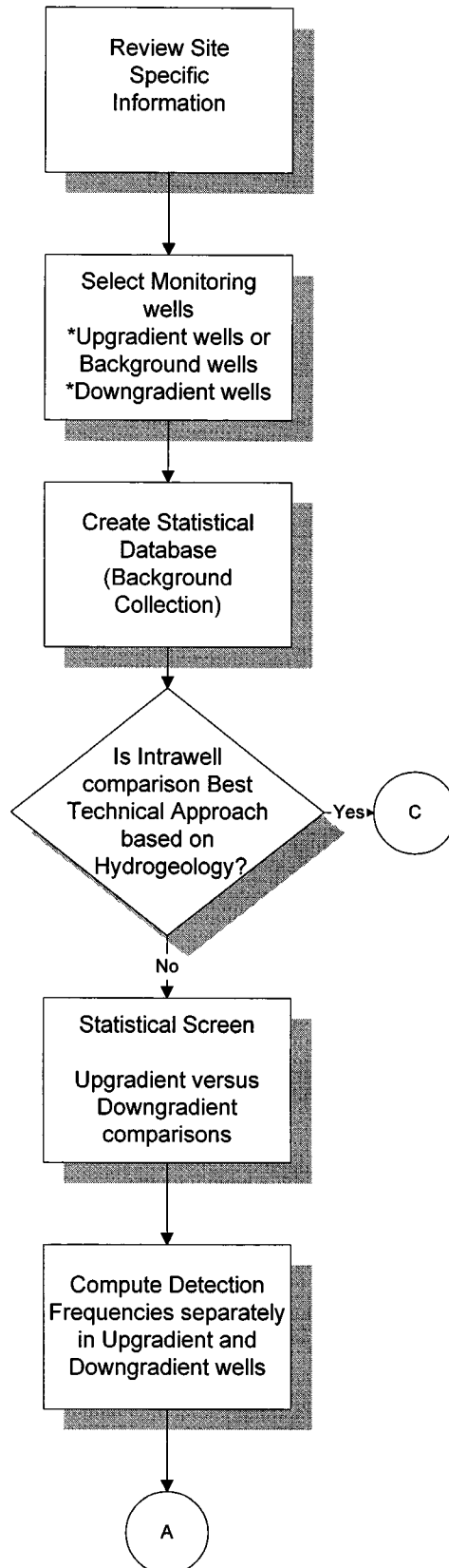


FIG. 1 Development of a Statistical Detection Monitoring Plan

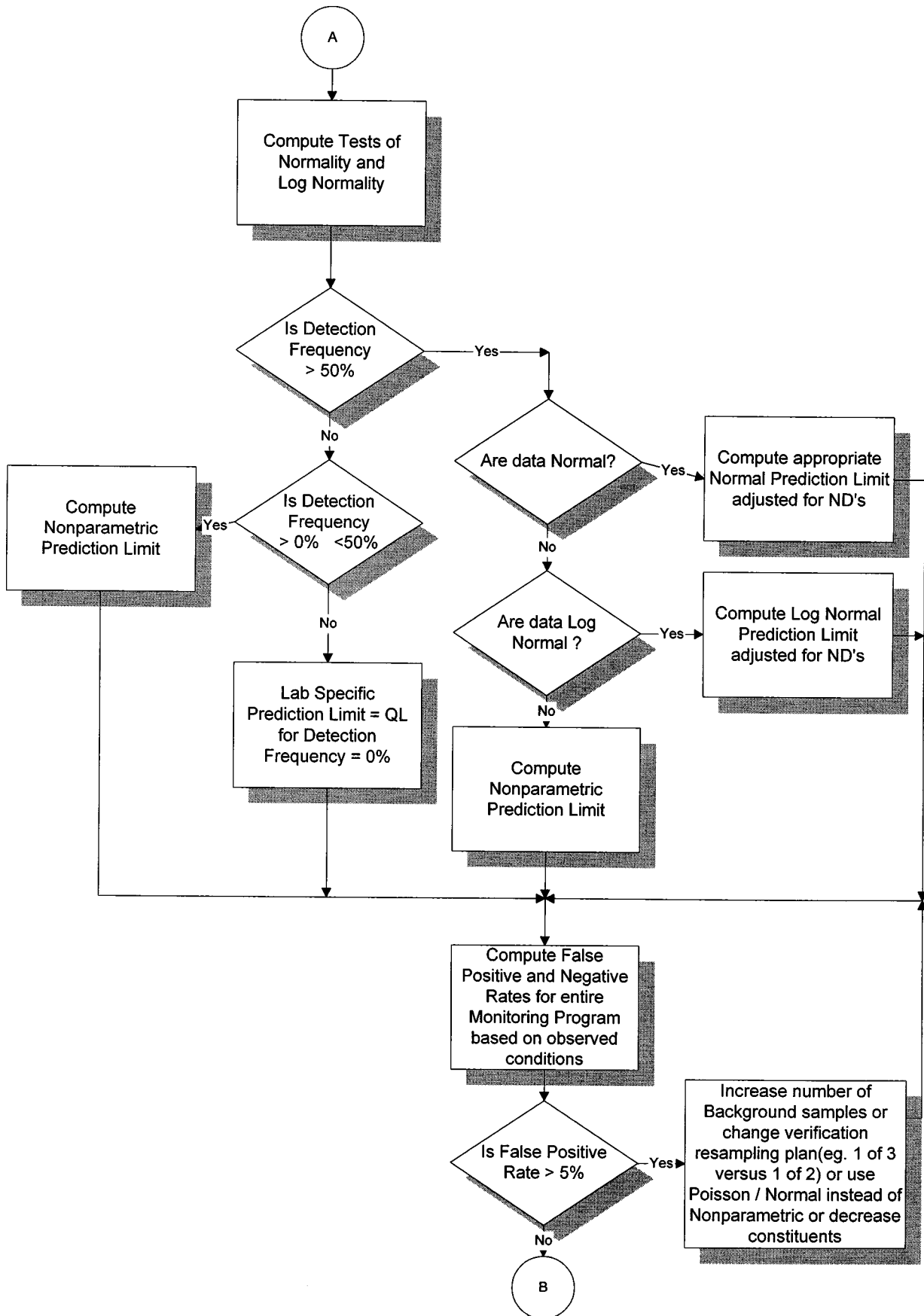


FIG. 1 (continued)

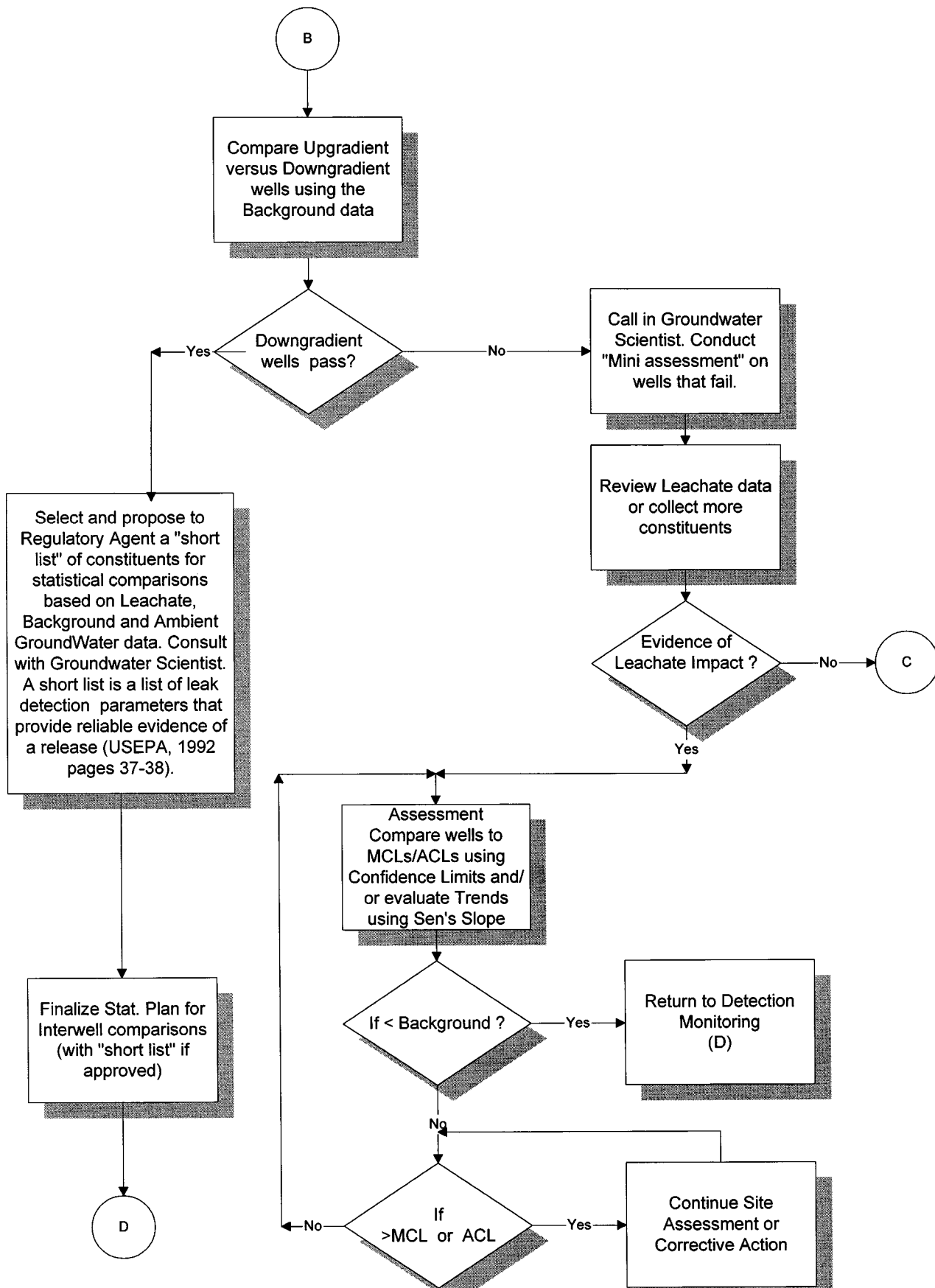


FIG. 1 (continued)



FIG. 1 (continued)

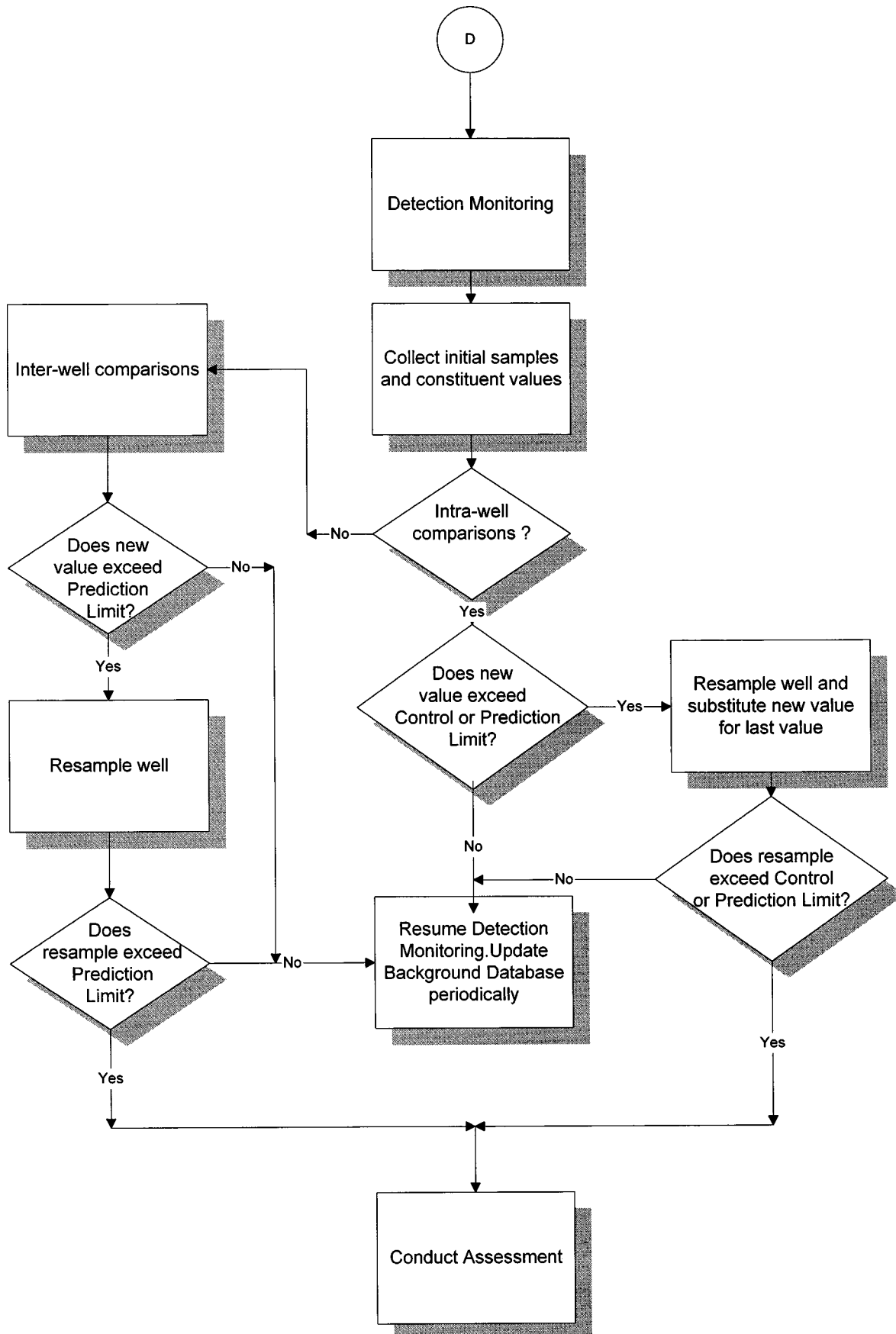


FIG. 1 (continued)

5. Procedure

NOTE 1—In the following, an overview of the general procedure is described with specific technical details described in Section 6.

5.1 Detection Monitoring:

5.1.1 Upgradient Versus Downgradient Comparisons:

5.1.1.1 Detection frequency $\geq 50\%$.

5.1.1.2 If the constituent is normally distributed, compute a normal prediction limit (5) selecting the false positive rate based on number of wells, constituents, and verification resamples (6) adjusting estimates of sample mean and variance for nondetects.

5.1.1.3 If the constituent is lognormally distributed, compute a lognormal prediction limit (7).

5.1.1.4 If the constituent is neither normally nor lognormally distributed, compute a nonparametric prediction limit (7) unless background is insufficient to achieve a 5% site-wide false positive rate. In this case, use a normal distribution until sufficient background data are available (7).

5.1.1.5 If the background detection frequency is greater than zero but less than 50%.

5.1.1.6 Compute a nonparametric prediction limit and determine if the background sample size will provide adequate protection from false positives.

5.1.1.7 If insufficient data exist to provide a site-wide false positive rate of 5%, more background data must be collected.

5.1.1.8 As an alternative to 5.1.1.7 use a Poisson prediction limit which can be computed from any available set of background measurements regardless of the detection frequency (see 2.2.4 of Ref (4)).

5.1.1.9 If the background detection frequency equals zero, use the laboratory-specific QL (recommended) or limits required by applicable regulatory agency (8).³

5.1.1.10 This only applies for those wells and constituents that have at least 13 background samples. Thirteen samples provides a 99% confidence nonparametric prediction limit with one resample for a single well and constituent (see Table 1).

5.1.1.11 If less than 13 samples are available more background data must be collected to use the nonparametric prediction limit.

5.1.1.12 An alternative would be to use a Poisson prediction limit that can be computed from four or more background measurements regardless of the detection frequency and can adjust for multiple wells and constituents.

5.1.1.13 If downgradient wells fail, determine cause.

5.1.1.14 If the downgradient wells fail because of natural or off-site causes, select constituents for intra-well comparisons (9).

5.1.1.15 If site impacts are found, a site plan for assessment monitoring may be necessary (10).

5.1.2 Intra-well Comparisons:

5.1.2.1 For those facilities that either have no definable hydraulic gradient, have no existing contamination, have too few background wells to meaningfully characterize spatial

variability (for example, a site with one upgradient well or a facility in which upgradient water quality is either inaccessible or not representative of downgradient water quality), compute intra-well comparisons using combined Shewart-CUSUM control charts (9).⁴

5.1.2.2 For those wells and constituents that fail upgradient versus downgradient comparisons, compute combined Shewart-CUSUM control charts. If no volatile organic compounds (VOCs) or hazardous metals are detected and no trend is detected in other indicator constituents, use intra-well comparisons for detection monitoring of those wells and constituents.

5.1.2.3 If data are all non-detects after 13 quarterly sampling events, use the QL as the nonparametric prediction limit (8). Thirteen samples provides a 99% confidence nonparametric prediction limit with one resample (1). Note that 99% confidence is equivalent to a 1% false positive rate, and pertains to a single comparison (that is, well and constituent) and not the site-wide error rate (that is, all wells and constituents) that is set to 5%.

5.1.2.4 If detection frequency is greater than zero (that is, the constituent is detected in at least one background sample) but less than 25%, use the nonparametric prediction limit that is the largest (or second largest) of at least 13 background samples.

5.1.2.5 As an alternative to 5.1.2.3 and 5.1.2.4 compute a Poisson prediction limit following collection of at least four background samples. Since the mean and variance of the Poisson distribution are the same, the Poisson prediction limit is defined even if there is no variability (for example, even if the constituent is never detected in background). In this case, one half of the quantification limit is used in place of the measurements, and the Poisson prediction limit can be computed directly.

5.1.3 Verification Resampling:

5.1.3.1 Verification resampling is an integral part of the statistical methodology (see Section 5 of Ref (4)). Without verification resampling much larger prediction limits would be required to obtain a site-wide false positive rate of 5%. The resulting false negative rate would be dramatically increased.

5.1.3.2 Verification resampling allows sequential application of a much smaller prediction limit, therefore minimizing both false positive and false negative rates.

5.1.3.3 A statistically significant exceedance is not declared and should not be reported until the results of the verification resample are known. The probability of an initial exceedance is much higher than 5% for the site as a whole.

5.1.3.4 Note that in the parametric case requiring passage of two verification resamples (for example, in the state of California regulation) will lead to higher false negative rates (for a fixed false positive rate) because larger prediction limits are required to achieve a site-wide false positive rate of 5% than for a single verification resample; hence, the preferred methods are pass one verification resample or pass one of two verification resamples. Also note that nonparametric limits requiring

³ Note, if background detection frequency is zero, one should question whether the analyte is a useful indicator of contamination. If it is not, statistical testing of the constituent should not be performed.

⁴ Some examples of inaccessible or nonrepresentative background upgradient wells may include slow moving ground water, radial or convergent flow, or sites that straddle ground-water divides.

TABLE 1 Probability That the First Sample or the Verification Resample Will Be Below the Maximum of n Background Measurements at Each of k Monitoring Wells for a Single Constituent

Previous n	Number of Monitoring Wells (k)														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
4	0.933	0.881	0.838	0.802	0.771	0.744	0.720	0.698	0.679	0.661	0.645	0.630	0.617	0.604	0.592
5	0.952	0.913	0.879	0.849	0.823	0.800	0.779	0.760	0.742	0.726	0.711	0.697	0.684	0.672	0.661
6	0.964	0.933	0.906	0.882	0.860	0.840	0.822	0.805	0.789	0.774	0.761	0.748	0.736	0.725	0.714
7	0.972	0.947	0.925	0.905	0.886	0.869	0.853	0.838	0.825	0.812	0.799	0.788	0.777	0.766	0.757
8	0.978	0.958	0.939	0.922	0.906	0.891	0.878	0.864	0.852	0.841	0.830	0.819	0.809	0.800	0.791
9	0.982	0.965	0.949	0.935	0.921	0.908	0.896	0.885	0.874	0.864	0.854	0.844	0.835	0.827	0.818
10	0.985	0.971	0.957	0.945	0.933	0.922	0.911	0.901	0.891	0.882	0.873	0.865	0.857	0.849	0.841
11	0.987	0.975	0.964	0.953	0.942	0.933	0.923	0.914	0.906	0.897	0.889	0.882	0.874	0.867	0.860
12	0.989	0.979	0.969	0.959	0.950	0.941	0.933	0.925	0.917	0.910	0.902	0.896	0.889	0.882	0.876
13	0.990	0.981	0.973	0.964	0.956	0.948	0.941	0.934	0.927	0.920	0.914	0.907	0.901	0.895	0.889
14	0.992	0.984	0.976	0.969	0.961	0.954	0.948	0.941	0.935	0.929	0.923	0.917	0.912	0.906	0.901
15	0.993	0.986	0.979	0.972	0.966	0.959	0.953	0.947	0.942	0.936	0.931	0.926	0.920	0.915	0.910
16	0.993	0.987	0.981	0.975	0.969	0.964	0.958	0.953	0.948	0.943	0.938	0.933	0.928	0.923	0.919
17	0.994	0.988	0.983	0.978	0.972	0.967	0.962	0.957	0.953	0.948	0.943	0.939	0.935	0.930	0.926
18	0.995	0.990	0.985	0.980	0.975	0.970	0.966	0.961	0.957	0.953	0.949	0.944	0.940	0.937	0.933
19	0.995	0.991	0.986	0.982	0.977	0.973	0.969	0.965	0.961	0.957	0.953	0.949	0.946	0.942	0.938
20	0.996	0.991	0.987	0.983	0.979	0.975	0.972	0.968	0.964	0.960	0.957	0.953	0.950	0.947	0.943
25	0.997	0.994	0.992	0.989	0.986	0.984	0.981	0.978	0.976	0.973	0.971	0.968	0.966	0.964	0.961
30	0.998	0.996	0.994	0.992	0.990	0.988	0.986	0.984	0.983	0.981	0.979	0.977	0.975	0.974	0.972
35	0.998	0.997	0.996	0.994	0.993	0.991	0.990	0.988	0.987	0.986	0.984	0.983	0.981	0.980	0.979
40	0.999	0.998	0.997	0.995	0.994	0.993	0.992	0.991	0.990	0.989	0.988	0.987	0.985	0.984	0.983
45	0.999	0.998	0.997	0.996	0.995	0.995	0.994	0.993	0.992	0.991	0.990	0.989	0.988	0.987	0.987
50	0.999	0.998	0.998	0.997	0.996	0.996	0.995	0.994	0.993	0.993	0.992	0.991	0.990	0.990	0.989
60	0.999	0.999	0.998	0.998	0.997	0.997	0.996	0.996	0.995	0.995	0.994	0.994	0.993	0.993	0.992
70	1.00	0.999	0.999	0.998	0.998	0.998	0.997	0.997	0.997	0.996	0.996	0.995	0.995	0.995	0.994
80	1.00	0.999	0.999	0.999	0.998	0.998	0.998	0.998	0.997	0.997	0.997	0.996	0.996	0.996	0.996
90	1.00	1.00	0.999	0.999	0.999	0.999	0.998	0.998	0.998	0.998	0.997	0.997	0.997	0.997	0.996
100	1.00	1.00	0.999	0.999	0.999	0.999	0.999	0.998	0.998	0.998	0.998	0.998	0.997	0.997	0.997

Previous n	Number of Monitoring Wells (k)														
	20	25	30	35	40	45	50	55	60	65	70	75	80	90	100
4	0.542	0.504	0.474	0.449	0.428	0.410	0.394	0.380	0.367	0.356	0.345	0.336	0.327	0.312	0.299
5	0.612	0.574	0.543	0.517	0.495	0.476	0.459	0.443	0.430	0.417	0.406	0.396	0.386	0.369	0.355
6	0.668	0.631	0.600	0.574	0.552	0.532	0.514	0.499	0.484	0.472	0.460	0.449	0.439	0.420	0.405
7	0.713	0.678	0.648	0.623	0.600	0.580	0.563	0.547	0.532	0.519	0.507	0.496	0.485	0.466	0.450
8	0.750	0.717	0.688	0.664	0.642	0.622	0.605	0.589	0.574	0.561	0.549	0.537	0.527	0.507	0.490
9	0.781	0.750	0.723	0.699	0.678	0.659	0.642	0.626	0.612	0.598	0.586	0.574	0.564	0.544	0.527
10	0.807	0.777	0.752	0.729	0.709	0.691	0.674	0.659	0.644	0.631	0.619	0.608	0.597	0.578	0.560
11	0.828	0.801	0.777	0.755	0.736	0.718	0.702	0.687	0.674	0.661	0.649	0.638	0.627	0.608	0.590
12	0.847	0.821	0.799	0.778	0.760	0.743	0.727	0.713	0.700	0.687	0.675	0.664	0.654	0.635	0.618
13	0.862	0.839	0.817	0.798	0.781	0.764	0.750	0.736	0.723	0.711	0.699	0.689	0.678	0.660	0.643
14	0.876	0.854	0.834	0.816	0.799	0.784	0.769	0.756	0.744	0.732	0.721	0.710	0.701	0.682	0.666
15	0.888	0.867	0.848	0.831	0.815	0.801	0.787	0.774	0.762	0.751	0.740	0.730	0.721	0.703	0.686
16	0.898	0.879	0.861	0.845	0.830	0.816	0.803	0.791	0.779	0.768	0.758	0.748	0.739	0.722	0.706
17	0.907	0.889	0.872	0.857	0.843	0.830	0.817	0.806	0.794	0.784	0.774	0.765	0.756	0.739	0.723
18	0.914	0.898	0.882	0.868	0.855	0.842	0.830	0.819	0.808	0.798	0.789	0.780	0.771	0.754	0.739
19	0.921	0.906	0.891	0.878	0.865	0.853	0.842	0.831	0.821	0.811	0.802	0.793	0.785	0.769	0.754
20	0.928	0.913	0.899	0.886	0.874	0.863	0.852	0.842	0.832	0.823	0.814	0.806	0.798	0.782	0.768
25	0.950	0.939	0.929	0.919	0.910	0.901	0.892	0.884	0.876	0.869	0.862	0.855	0.848	0.835	0.823
30	0.963	0.955	0.947	0.940	0.932	0.925	0.919	0.912	0.906	0.900	0.894	0.888	0.882	0.872	0.861
35	0.972	0.966	0.959	0.954	0.948	0.942	0.937	0.931	0.926	0.921	0.916	0.911	0.907	0.898	0.889
40	0.978	0.973	0.968	0.963	0.958	0.954	0.949	0.945	0.941	0.936	0.932	0.928	0.924	0.917	0.909
45	0.982	0.978	0.974	0.970	0.966	0.962	0.959	0.955	0.951	0.948	0.944	0.941	0.938	0.931	0.925
50	0.985	0.982	0.979	0.975	0.972	0.969	0.966	0.963	0.959	0.956	0.954	0.951	0.948	0.942	0.937
60	0.990	0.987	0.985	0.982	0.980	0.978	0.975	0.973	0.971	0.968	0.966	0.964	0.962	0.958	0.954
70	0.992	0.990	0.989	0.987	0.985	0.983	0.981	0.980	0.978	0.976	0.974	0.973	0.971	0.968	0.965
80	0.994	0.993	0.991	0.990	0.988	0.987	0.986	0.984	0.983	0.981	0.980	0.979	0.977	0.975	0.972
90	0.995	0.994	0.993	0.992	0.991	0.990	0.988	0.987	0.986	0.985	0.984	0.983	0.982	0.980	0.978
100	0.996	0.995	0.994	0.993	0.992	0.991	0.991	0.990	0.989	0.988	0.987	0.986	0.985	0.983	0.982

passage of two verification resamples will result in need for a larger number of background samples than are typically available (see 6.3.3.1) (1).

5.1.4 False Positive and False Negative Rates:

5.1.4.1 Conduct simulation study based on current monitoring network, constituents, detection frequencies, and distributional form of each monitoring constituent (see Appendix B of Ref (4)). The specific objectives of the simulation study are to determine if the false positive and false negative rates of the

current monitoring program as a whole are acceptable and to determine if changes in verification resampling plans or choice of nonparametric versus Poisson prediction limits or inter-well versus intra-well comparison strategies will improve the overall performance of the detection monitoring program.

5.1.4.2 Project frequency of which verification resamples will be required and false assessments for site as a whole for each monitoring event based on the results of the simulation study. In this way the owner/operator will be able to anticipate

the required amount of future sampling.

5.1.4.3 As a general guideline, a site-wide false positive rate of 5 % and a false negative rate of approximately 5 % for differences on the order of three to four standard deviation units are recommended. Note that USEPA recommends simulating the most conservative case of a release that effects a single constituent in a single downgradient well. In practice, multiple constituents in multiple wells will be impacted, therefore, the actual false negative rates may be considerably smaller than estimates obtained by means of simulation.

5.1.5 *Use of DLs and QLs in Ground-Water Monitoring:*

5.1.5.1 The DLs indicate that the analyte is present in the sample with confidence.

5.1.5.2 The QLs indicate that the true quantitative value of the analyte is close to the measured value.

5.1.5.3 For analytes with estimated concentration exceeding the DL but not the QL, it can be concluded that the true concentration is greater than zero; however, uncertainty in the instrument response is by definition too large to make a reliable quantitative determination. Note that in a qualitative sense, values between the DL and QL are greater than values below the DL, and this rank ordering can be used in a nonparametric method.

5.1.5.4 If the laboratory-specific DL for a given compound is 3 μ g/L, and the QL for the same compound is 6 μ g/L, then a detection of that compound at 4 μ g/L could actually represent a true concentration of anywhere between 0 and 6 μ g/L. The true concentration may well be less than the DL **(1,2,11)**.

5.1.5.5 Direct comparison of a single value to a maximum concentration level (MCL), or any other concentration limit, is not adequate to demonstrate noncompliance unless the concentration is larger than the QL.

5.1.5.6 Verification resampling applies to this case as well.

6. Report

6.1 This section provides a description of the specific statistical methods referred to in this guide. Note that specific recommendations for any given facility require an interdisciplinary site-specific study that encompasses knowledge of the facility, its hydrogeology, geochemistry, and study of the false positive and false negative error rates that will result. Performing a correct statistical analysis, such as nonparametric prediction limits, in the wrong situation (for example, when there are too few background measurements) can lead to erroneous conclusions.

6.2 *Upgradient Versus Downgradient Comparisons:*

6.2.1 *Case One—Compounds Quantified in All Background Samples:*

6.2.1.1 Test normality of distribution using the multiple group version of the Shapiro-Wilk test applied to n background measurements **(12)**. The multiple group version of the Shapiro-Wilk test takes into consideration that background measurements are nested within different background monitoring wells, hence the original Shapiro-Wilk test does not directly apply.

NOTE 2—Background wells used for inter-well comparisons may in some cases include wells that are not hydraulically upgradient of the site.

6.2.1.2 Alternatively, residuals from the mean of each upgradient well can be pooled together and tested using the

single group version of the Shapiro-Wilk test **(13)**.

6.2.1.3 The need for a multiple group test to incorporate spatial variability among upgradient wells also raises the question of validity of upgradient versus downgradient comparisons. Where significant spatial variability exists, it may not be possible to obtain a representative upgradient background, and intra-well comparisons may be required. A one-way analysis of variance (ANOVA) applied to the upgradient well data provides a good way of testing for significant spatial variability.

6.2.1.4 If normality is not rejected, compute the 95 % prediction limit as follows:

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

where:

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

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} \quad (3)$$

- α = false positive rate for each individual test,
 $t_{[n-1,\alpha]}$ = one-sided $(1 - \alpha)$ 100 % point of Student's t distribution on $n - 1$ df, and
 n = number of background measurements. Select α as the minimum of 0.01 or one of the following:

(1) Pass the first or one of one verification resample:

$$\alpha = (1 - 0.95^{1/k})^{1/2} \quad (4)$$

(2) Pass the first or one of two verification resamples:

$$\alpha = (1 - 0.95^{1/k})^{1/3} \quad (5)$$

(3) Pass the first or two of two verification resamples:

$$\alpha = \sqrt{1 - 0.95^{1/k}} \sqrt{1/2} \quad (6)$$

where:

k = number of comparisons (that is, monitoring wells times constituents (see section 5.2.2 of Ref **(4)**).

6.2.1.5 Note that these formulas for computing the adjusted individual comparison α all ignore two sources of dependence: comparisons for a given constituent are all made against the same background and concentrations of the indicator constituents may be positively correlated over time. Solution of the first problem has been provided by Refs **(1)** and **(14)** and has provided detailed tabulation of factors that can be used in computing the exact prediction limits. In terms of the second problem, constituents that are highly correlated (based on pairwise correlations) could be eliminated, not from the statistical analysis, but from the total set of comparisons used to compute α , leading to more powerful and realistic prediction limits.

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

6.2.1.7 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 follows:

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

where:

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

and:

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

6.2.1.8 If log transformation does not bring about normality (that is, the probability of G is less than 0.01), compute nonparametric prediction limits (Option—Compute normal prediction limit).

6.2.2 *Case Two—Compounds Quantified in at Least 50 % of All Background Samples:*

6.2.2.1 Apply the multiple group Shapiro-Wilk test to the n_1 quantified measurements only.

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

$$\bar{x} = \left(1 - \frac{n_0}{n}\right) \bar{x}' \quad (10)$$

where:

\bar{x}' = average of the n_1 detected values, and

n_0 = number of samples in which the compound is not detected. The standard deviation is:

$$s = \sqrt{\left(1 - \frac{n_0}{n}\right) s'^2 + \frac{n_0}{n} \left(1 - \frac{n_0 - 1}{n - 1}\right) \bar{x}'^2} \quad (11)$$

where s' is the standard deviation of the n_1 detected measurements. The normal prediction limit can then be computed as previously described. This method is due to Aitchison (see 2.2.2 of Ref (4) and (15)). Note that this method imputes nondetects as zero concentrations.

6.2.2.3 A good alternative to Aitchison's method is Cohen's maximum likelihood estimator (16). Extensive tables and computational details are also provided in Gibbons, 1991. A useful approach to selecting between the two methods is described in 2.2.1 of Ref (4).

6.2.2.4 If the multiple group Shapiro-Wilk test reveals that the data are lognormally distributed, replace \bar{x}' with \bar{y}' and s' and s'_y in the equations for \bar{x} and s . The lognormal prediction limit may then be computed as previously described.

NOTE 3—This adjustment only applies to positive random variables. The natural logarithm of concentration less than 1 are negative and therefore the adjustment does not apply. For this reason we add 1 to each value (for example, $\log_e(x_i + 1) \geq 0$), compute the prediction limit on a log scale and then subtract one from the antilog of the prediction limit.

6.2.2.5 If the data are neither normally or lognormally distributed, compute a nonparametric prediction limit. (Option—compute normal prediction limit).

6.2.3 *Case Three—Compounds Quantified in Less Than 50 % of All Background Samples:*

6.2.3.1 In this application, the nonparametric prediction limit is the largest concentration found in n upgradient measurements (see section 4.2.1 of Ref (8)).

6.2.3.2 Gibbons (18,19) has shown that the confidence associated with this decision rule, following one or more

verification resamples, is a function of the multivariate extension of the hypergeometric distribution (see section 5.2.3 of Ref (8)).

6.2.3.3 Complete tabulations of confidence levels for $n = 4, \dots, 100$, $k = 1, \dots, 100$ future comparisons (for example, monitoring wells), and a variety of verification resampling plans are presented in (1). For example with five monitoring wells and ten constituents (that is, 50 comparisons), 40 background measurements would be required to provide 95 % confidence (see section 5.2.3 of Ref (4)). Table 1 displays confidence levels for a single verification resample.

6.2.3.4 As an option to the nonparametric prediction limits, compute Poisson prediction limits. Poisson prediction limits are useful for those cases in which there are too few background measurements to achieve an adequate site-wide false positive rate using the nonparametric approach. Gibbons (19) derived the original Poisson prediction limit. Cameron (20) found that use of a normal multiplier in place of Student's t -distribution resulted in a more powerful test, thus the Poisson prediction limit is:

$$\text{Poisson PL} = y/n + \frac{z^2}{2n} + z/n \sqrt{y(1+n) + z^2/4} \quad (12)$$

where y is the sum of the detected measurements or the quantification limit for those samples in which the constituent was not detected, and z is the $(1 - \alpha)$ 100 upper percentage point of the normal distribution, where α is computed as in 6.2.1.4.

NOTE 4—If the Poisson prediction unit is less than the quantification limit, recompute the prediction limit substituting the quantification limit for the nondetects.

6.3 *Intra-Well Comparisons:*

6.3.1 One particularly good method for computing intra-well comparisons is the combined Shewart-CUSUM control chart (see 6.1 in Ref (4)). The method is sensitive to both gradual and rapid releases and is also useful as a method of detecting " trends" in data. Note that this method should be used on wells unaffected by the landfill. There are several approaches to implementing the method, and in the following, one useful way is described as well as discussion of some statistical properties.

6.3.2 *Assumptions:*

6.3.2.1 The combined Shewart-CUSUM control chart procedure assumes that the data are independent and normally distributed with a fixed mean μ and constant variance σ^2 . The most important assumption is independence, and as a result, wells should be sampled no more frequently than quarterly. In some cases, where ground water moves relatively quickly, it may be possible to accelerate background sampling to eight samples in a single year; however, this should only be done to establish background and not for routine monitoring. The assumption of normality is somewhat less of a concern, and if problematic, natural log or square root transformation of the observed data should be adequate for most practical applications. For this method, nondetects can be replaced by the quantification limit without serious consequence. This procedure should only be applied to those constituents that are

detected at least in 25 % of all samples, otherwise, σ^2 is not adequately defined.

6.3.2.2 When large intra-well background databases are available, (for example, three years or more of at least semiannual monitoring) obvious cyclic or trend patterns can be removed from both the baseline data and from the future data to be plotted on the chart. Similarly, when the background database consists of eight or more background measurements, use of Aitchison's (15) or Cohen's (16) methods for computing the background mean and standard deviation can be used in place of simple imputation of the quantification limit.

6.3.3 Nondetects:

6.3.3.1 For those well and constituent combinations in which the detection frequency is less than 25 %, the data should be displayed graphically until a sufficient number of measurements are available to provide 99 % confidence (that is, 1 % false positive rate) for an individual well and constituent using a nonparametric prediction limit, which in this context is the maximum detected value out of the n historical measurements. As previously discussed this amounts to 13 background samples for 1 resample, 8 background samples for pass 1 of 2 resamples and 18 background samples for pass 2 of 2 resamples. If nonparametric prediction limits are to be used for intra-well comparisons of rarely detected constituents, 2 verification resamples will often be required, and failure will only be indicated if both measurements exceed the limit (that is, the maximum of the first 8 samples).

6.3.3.2 Note that these background sample sizes provide 99 % confidence for a single future comparison and not all of the wells and constituents for which they will actually be applied. Adjustment for multiple comparisons will require even larger background sample sizes that may not be possible to obtain at most facilities. In light of this, the recommendations in 6.3.3.1 provide a minimum requirement.

6.3.3.3 For those cases in which the detection frequency is greater than 25 %, substitute the QL (or where there are multiple QLs, the median QL) for the nondetects. In this way, changes in quantification limits do not appear to be significant trends.

6.3.3.4 If nothing is detected in 8, 13, or 18 independent samples (depending on resampling strategy), use the quantification limit as the nonparametric prediction limit.

6.3.3.5 As in the previously described inter-well comparisons, optional use of Poisson prediction limits as an alternative to nonparametric prediction limits for rarely detected constituents (that is, less than 25 % detects) is recommended when the number of background measurements is small. Poisson prediction limits can be computed after eight background measurements regardless of detection frequency.

6.3.4 Procedure:

6.3.4.1 Require that at least eight historical independent samples are available to provide reliable estimates of the mean μ and standard deviation σ , of the constituent's concentration in each well.

6.3.4.2 Select the three Shewart-CUSUM parameters, h , (the value against which the cumulative sum will be compared), c (a parameter related to the displacement that should be quickly detected), and SCL (the upper Shewart limit that is

the number of standard deviation units for an immediate release). Lucas (21) and Starks (22) suggest that $c = 1$, $h = 5$, and $SCL = 4.5$ are most appropriate for ground-water monitoring applications. This sentiment is echoed by USEPA in their interim final guidance document (23).

6.3.4.3 Denote the new measurement at time-point t_i as x_i and compute the standardized value z_i :

$$z_i = \frac{x_i - x}{s} \quad (13)$$

where x and s are the mean and standard deviation of at least eight historical measurements for that well and constituent (collected in a period of no less than one year).

6.3.4.4 At each time period, t_i , compute the cumulative sum S_i , as:

$$S_i = \max [0, (z_i - c) + S_{i-1}] \quad (14)$$

where: $\max[A, B]$ is the maximum of A and B , starting with $S_0 = 0$.

6.3.4.5 Plot the values of S_i (y -axis) versus t_i (x -axis) on a time chart. Declare an "out-of-control" situation on sampling period t_i if for the first time, $S_i \geq h$ or $z_i \geq SCL$. Any such designation, however, must be verified on the next round of sampling, before further investigation is indicated.

6.3.4.6 The reader should note that unlike prediction limits that provide a fixed confidence level (for example, 95 %) for a given number of future comparisons, control charts do not provide explicit confidence levels, and do not adjust for the number of future comparisons. The selection of $h = 5$, $SCL = 4.5$ and $c = 1$ is based on USEPA's own review of the literature and simulations 21,22, and 23). The USEPA indicates that these values "allow a displacement of two standard deviations to be detected quickly." Since 1.96 standard deviation units corresponds to 95 % confidence on a normal distribution, we can have approximately 95 % confidence for this test method as well. In practice, setting $h = SCL = 4.5$ results in a single limit with no compromise in leak detection capabilities.

6.3.4.7 In terms of plotting the results, it is more intuitive to plot values in their original metric (for example, microgram per litre) rather than in standard deviation units. In this case, $h = SCL = \bar{x} + 4.5s$, and the S_i are converted to the concentration metric by the transformation $S_i^* s + \bar{x}$, noting that when normalized (that is, in standard deviation units) $\bar{x} = 0$ and $s = 1$ so that $h = SCL = 4.5$ and $S_i^* 1 + 0 = S_i$. Note that when $n \geq 12$ recompute the mean and standard deviation and adjust the control limits $h = SCL = 4.0$ and $c = 0.75$.

6.3.5 Outliers:

6.3.5.1 From time to time, inconsistently large or small values (outliers) can be observed due to sampling, laboratory, transportation, transcription errors, or even by chance alone. Verification resampling will tremendously reduce the probability of concluding that an impact has occurred if such an anomalous value is obtained for any of these reasons. However, nothing has eliminated the chance that such errors might be included in the historical measurements for a particular well and constituent. If such erroneous values (either too high or too low) are included in the historical database, the result would be

an artificial increase in the magnitude of the control limit, and a corresponding increase in the false negative rate of the statistical test (that is, conclude that there is no site impact when in fact there is).

6.3.5.2 To remove the possibility of this type of error, the historical data are screened for each well and constituent for the existence of outliers (see 6.2 in Ref (4)) using the well-known method described by Dixon (24). These outlying data points are indicated on the control charts (using a different symbol), but are excluded from the measurements that are used to compute the background mean and standard deviation. In the future, new measurements that turn out to be outliers, in that they exceed the control limit, will be dealt with by verification resampling in downgradient wells only.

6.3.5.3 This same outlier detection algorithm is applied to each upgradient well and constituent to screen outliers for inter-well comparisons as well.

6.3.6 Existing Trends:

6.3.6.1 If contamination is preexisting, trends will often be observed in the background database from which the mean and variance are computed. This will lead to upward biased estimates and grossly inflated control limits. To remove this possibility, first screen the background data for each well and constituent for trend using Sen's nonparametric estimate of trend (25). Confidence limits for this trend estimate are given by Gilbert (26). A significant trend is one in which the 99 % lower confidence bound is greater than zero. In this way, even preexisting trends in the background dataset will be detected.

6.3.6.2 When significant trends in background are found, their source must be identified prior to continuation of detection monitoring since they may be evidence of a prior site impact. If the source of the trend is found to be unrelated to the facility, then an alternative indicator constituent may be required for that well or all wells at the facility.

6.3.7 Note on Verification Sampling:

6.3.7.1 It should be noted that when a new monitoring value is an outlier, perhaps due to a transcription error, sampling error, or analytical error, the Shewart and CUSUM portions of the control chart are affected quite differently. The Shewart portion of the control chart compares each individual new measurement to the control limit, therefore, the next monitoring event measurement constitutes an independent verification of the original result. In contrast, however, the CUSUM procedure incorporates all historical values in the computation, therefore, the effect of the outlier will be present for both the initial and verification sample: hence the statistical test will be invalid.

6.3.7.2 For example, assume $\bar{x} = 50$ and $s = 10$. On Quarter 1 the new monitoring value is 50, so $z = (50 - 50)/10 = 0$ and $S_i = \max[0, (z - 1) + 0] = 0$. On Quarter 2, a sampling error occurs (that is, documented as an error after review of chain of custody) and the reported value is 200, yielding $z = (200 - 50)/10 = 15$ and $S_i = \max[0, (15 - 1) + 0] = 14$, that is considerably larger than 4.5; hence an initial exceedance is recorded. On the next round of sampling, the previous result is not confirmed, because the result is back to 50. Inspection of the CUSUM, however, yields $z = (50 - 50)/10 = 0$ and $S_i = \max[0, (0 - 1) + 14] = 13$, that

would be taken as a confirmation of the exceedance, when in fact, no such confirmation was observed. For this reason, the verification must replace the suspected result in order to have an unbiased confirmation.

6.3.8 *Updating the Control Chart*—As monitoring continues and the process is shown to be in control, the background mean and variance should be updated periodically to incorporate these new data. Every year or two, all new data that are in control should be pooled with the initial samples and \bar{x} and s recomputed. These new values of \bar{x} and s will then be used in constructing future control charts. This updating process should continue for the life of the facility or monitoring program, or both (see 6.1 in Ref (8)).

6.3.9 *An Alternative Based on Prediction Limits*—An alternative approach to intra-well comparisons involves computation of well-specific prediction limits. Prediction limits are somewhat more sensitive to immediate releases but less sensitive to gradual releases than the combined Shewart-CUSUM control charts. Prediction limits are also less robust to deviations from distributional assumptions (1).

7. Restriction of Background Samples

7.1 Certain states have interpreted the regulations as indicating that background be confined to the first four samples collected in a day or a semiannual monitoring event or a year. This conflicts with federal regulation and guidance. The first approach (that is, four samples in a day) violates the assumption of independence and confounds day to day temporal and seasonal variability with potential contamination. As an analogy, consider setting limits on yearly ambient temperatures in Chicago by taking four temperature readings on July 4th. On that day the temperature varied between 78 and 82°F (26 and 28°C) yielding a prediction interval from 70 to 90°F (21 to 32°C). In January, the temperature in Chicago can be -20°F (-28°C). Clearly, in this example restriction of background leads to nonrepresentative prediction of future measurements. In the second approach restricting establishment of background to the first four events taken in six months underestimates the component of seasonal variability and can lead to elevated false positive or false negative rates. The net result is that comparisons of background water quality in the summer may not be representative of downgradient ground-water quality in the winter (for example, disposal of road salts increasing specific conductivity in the winter). In the third approach in which background is restricted to the first four quarterly measurements, independence is typically not an issue and background versus point of compliance monitoring well comparisons are not confounded with season for that year, however, background from this year may not reflect temporal variability in future years (for example, a drought condition). In addition, as previously pointed out in the temperature illustration, restriction of background to only four samples dramatically increases the size of the statistical prediction limit thereby increasing the false negative rate of the test (that is, the prediction limit is over five standard deviation units above the background mean concentration). The reason for this is that the uncertainty in the true mean concentration covers the majority of the normal distribution. As such, virtually any mean and standard deviation could be obtained by chance alone. If by

chance the values are low, false positive results will occur. If by chance the values are high, false negative results will occur. By increasing the background sample size, uncertainty in the sample-based mean and standard deviation decrease as does the size of the prediction limit, therefore both false positive and false negative rates are minimized.

7.2 In light of these considerations, it is always in the best interest to have the largest available background database consisting of independent and representative measurements. Two possible strategies used to obtain a larger background

database are add background wells to the monitoring system (this also facilitates characterization of spatial variability) and update the background database at appropriate intervals (that is, either continuously for inter-well or every year or two for intra-well) with new measurements that are determined to belong to the same background population.

8. Keywords

8.1 control charts; detection monitoring; ground water; prediction limits; statistics; waste disposal facilities

REFERENCES

- (1) Gibbons, R. D., *Statistical Methods for Ground-Water Monitoring*, John Wiley & Sons, 1994.
- (2) Currie, *Analytical Chemistry*, 40, 1968, pp. 586–593.
- (3) Koorse, *Environmental Law Reporter*, 19, 1989, pp. 10211–10222.
- (4) USEPA, “Addendum to Interim Final Guidance Document,” *Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities*, July, 1992.
- (5) USEPA 40CFR Part 264: “Statistical Methods for Evaluating Ground-Water Monitoring from Hazardous Waste Facilities; Final Rule,” *Federal Register*, 53, 196, 1988, pp. 39720–39731.
- (6) USEPA 40CFR 258.53(h)(2).
- (7) USEPA 40CFR 258.53(h)(1).
- (8) USEPA 40CFR 258.53(h)(5).
- (9) USEPA 40CFR 258.53(h)(3).
- (10) USEPA 40CFR 258.55.
- (11) Hubaux, A., and Vos, G., *Analytical Chemistry*, 42, 1970, pp. 849–855.
- (12) Wilk, M. B., and Shapiro, S. S., *Technometrics*, 10, No. 4, 1968, pp. 825–839.
- (13) Shapiro, S. S., and Wilk, M. B., *Biometrika*, 52, 1965, pp. 591–611.
- (14) Davis, C. B. and McNichols, R. J., *Technometrics*, 29, 1987, pp. 359–370.
- (15) Aitchison, J., *Journal of the American Statistical Association*, 50, 1955, pp. 901–908.
- (16) Cohen, A. C., *Technometrics*, 3, 1961, pp. 535–541.
- (17) Gibbons, R. D., *Ground Water*, 28, 1990, pp. 235–243.
- (18) Gibbons, R. D., *Ground Water*, 29, 1991, pp. 729–736.
- (19) Gibbons, R. D., *Ground Water*, 25, 1987, pp. 572–580.
- (20) Cameron, K., (1995) EPA/530-R-93-003.
- (21) Lucas, J. M., *Journal of Quality Technology*, 14, 1982, pp. 51–59.
- (22) Starks, T. H., “Evaluation of Control Chart Methodologies for RCRA Waste Sites,” *USEPA Technical Report CR814342-01-3*, 1988.
- (23) *Statistical Analysis of Ground-Water Monitoring Data at RCRA Facilities*, 1989.
- (24) *Biometrics*, 9, 1953, pp. 74–89.
- (25) Sen, P. K., *Journal of the American Statistical Association*, 63, 1968, pp. 1379–1389.
- (26) Gilbert, R. O., *Statistical Methods for Environmental Pollution Monitoring*, Van Nostrand Reinhold, New York, 1987.

The American Society for Testing and Materials takes no position respecting the validity of any patent rights asserted in connection with any item mentioned in this standard. Users of this standard are expressly advised that determination of the validity of any such patent rights, and the risk of infringement of such rights, are entirely their own responsibility.

This standard is subject to revision at any time by the responsible technical committee and must be reviewed every five years and if not revised, either reapproved or withdrawn. Your comments are invited either for revision of this standard or for additional standards and should be addressed to ASTM Headquarters. Your comments will receive careful consideration at a meeting of the responsible technical committee, which you may attend. If you feel that your comments have not received a fair hearing you should make your views known to the ASTM Committee on Standards, 100 Barr Harbor Drive, West Conshohocken, PA 19428.