Designation: E3242 − 23

Standard Guide for Developing Representative Background Concentrations at Sediment Sites—Data Evaluation and Development Methodologies[1]

This standard is issued under the fixed designation E3242; 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 describes data visualization, statistical, forensic chemistry and geochemical methodologies (including case studies) used in the evaluation of candidate background data sets; this evaluation leads to the development of representative background data sets for the sediment site. Statistical methodologies can then be applied to the representative background data sets to develop background threshold values (BTVs) that are measures of the upper limit of representative sediment background concentrations for the sediment site. In addition, representative background data sets and sediment site data sets can be compared using two-sample statistical tests to determine if there are statistically significant differences (at a specified confidence level) between the two data sets (such as, the median or mean values of the two data sets are significantly different).

1.1.1 This guide is intended to inform, complement, and support, but not supersede the guidelines established by local, state, tribal, federal, or international agencies.

1.2 Technically defensible representative sediment background concentrations are critical for several purposes (**1**).[2] These include sediment site delineation, establishing remedial goals and cleanup levels, remedy selection, assessment of risks posed by representative background concentrations, and establishing appropriate post-remedial monitoring plans.

1.3 The overarching framework for the development of representative sediment background concentrations at sediment sites is presented in Guide E3382. Guide E3240 provides a general discussion of how conceptual site model (CSM) development fits into the risk-based corrective action framework for contaminated sediment sites, while Guide E3382 provides a detailed discussion of the elements of a sediment site CSM that need to be considered when developing representative sediment background concentrations. Guide E3344 describes how to select an appropriate background reference area(s) from which to collect sediment samples for laboratory analysis. Guide E3164 describes the sampling methodologies to obtain sediment samples in the field (whether from the sediment site or background reference area[s]), while Guide E3163 discusses appropriate laboratory methodologies for the chemical analysis of potential contaminants of concern (PCOCs) in the sediment samples. Relevant content contained in Guides E3344 and E3382 is summarized herein, but the individual guides should be consulted for more detailed coverage of these topics.

1.4 This guide focuses on the approach for the development of representative sediment background concentrations used for remedial actions performed under various regulatory programs, including the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA). Although many of the references cited in this guide are CERCLA oriented, the guide is applicable to remedial actions performed under local, state, tribal, federal, and international cleanup programs. However, the guide does not describe requirements for each jurisdiction. The requirements for the regulatory entity under which the cleanup is being performed should be reviewed to confirm compliance.

1.5 This guide is designed to apply to contaminated sediment sites where sediment data have been collected and are readily available. Additionally, this guide assumes that risk assessments have been performed, so that the contaminants of concern (COCs) that exceed risk-based thresholds have been identified.

1.5.1 Furthermore, this guide presumes that the identified risk-based thresholds are low enough to pose corrective action implementation challenges, or the site is subject to recontamination from uncontrolled ongoing anthropogenic or natural sources, or both. In all cases, representative sediment background concentrations will be useful for determining the extent of corrective remedial actions (when used as remedial goals or cleanup levels), evaluating risks posed by representative background concentrations, and establishing appropriate post-remedial monitoring plans.

[1] This guide is under the jurisdiction of ASTM Committee E50 on Environmental Assessment, Risk Management and Corrective Action and is the direct responsibility of Subcommittee E50.04 on Corrective Action.

Current edition approved Dec. 1, 2023. Published January 2024. Originally approved in 2020. Last previous edition approved in 2020 as E3242–20. DOI: 10.1520/E3242–23

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

1.6 *Units*—The values stated in SI or CGS units are to be regarded as standard. No other units of measurement are included in this standard.

1.7 *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, health, and environmental practices and determine the applicability of regulatory limitations prior to use.*

1.8 *This international standard was developed in accordance with internationally recognized principles on standardization established in the Decision on Principles for the Development of International Standards, Guides and Recommendations issued by the World Trade Organization Technical Barriers to Trade (TBT) Committee.*

2. Referenced Documents

2.1 *ASTM Standards:*[3]

D6312 Guide for Developing Appropriate Statistical Approaches for Groundwater Detection Monitoring Programs at Waste Disposal Facilities
D7048 Guide for Applying Statistical Methods for Assessment and Corrective Action Environmental Monitoring Programs
D7659 Guide for Strategies for Surface Sampling of Metals and Metalloids for Worker Protection
D7720 Guide for Statistically Evaluating Measurand Alarm Limits when Using Oil Analysis to Monitor Equipment and Oil for Fitness and Contamination
E178 Practice for Dealing With Outlying Observations
E456 Terminology Relating to Quality and Statistics
E1689 Guide for Developing Conceptual Site Models for Contaminated Sites
E3163 Guide for Selection and Application of Analytical Methods and Procedures Used during Sediment Corrective Action
E3164 Guide for Contaminated Sediment Site Risk-Based Corrective Action – Baseline, Remedy Implementation and Post-Remedy Monitoring Programs
E3240 Guide for Risk-Based Corrective Action for Contaminated Sediment Sites
E3248 Guide for NAPL Mobility and Migration in Sediment – Conceptual Models for Emplacement and Advection
E3344 Guide for Developing Representative Sediment Background Concentrations at Sediment Sites—Selection of Background Reference Areas
E3382 Guide for Developing Representative Background Concentrations at Sediment Sites — Framework Overview, Including Conceptual Site Model Considerations

3. Terminology

3.1 *Definitions:*

3.1.1 *anthropogenic background, n*—human-made substances present in the environment due to human activities, not specifically related to current or historical site-related releases or activities. **E3344**

3.1.1.1 *Discussion*—The definition of "anthropogenic background" varies with jurisdiction. In some jurisdictions, the regulator defines anthropogenic background as having both human-made and naturally occurring components. (**2**)

3.1.2 *background (aka "reference"), n*—a term applied to substances, conditions, or locations that are similar to those found at a sediment site but not influenced by current or historical releases or activities from the sediment site; these are usually a combination of naturally occurring (consistently present in the environment but not influenced by human activity) and anthropogenic (influenced by human activity but not related to specific current or historical releases or activities at the sediment site) components. **E3382**

3.1.3 *candidate background data set, n*—a raw (that is, unprocessed) background data set obtained either by the collection of data from a background reference area(s), or by the extraction of background data from the sediment site data set, or a combination of both. **E3382**

3.1.3.1 *Discussion*—The candidate background data set must first be evaluated using the steps described in this guide to obtain a representative background data set.

3.1.4 *contaminant of concern (COC), n*—substances identified as posing a risk based on a tiered risk assessment and that may warrant corrective action. **E3382**

3.1.4.1 *Discussion*—Typically, all potential contaminants of concern (PCOCs) identified for a sediment site are evaluated in the risk assessment process. PCOCs that have sediment concentrations greater than risk-based thresholds identified in the risk assessment process are defined as COCs. Thus, the COCs identified for a sediment site are a subset of the PCOCs identified for that site.

3.1.5 *distribution, n—as used in statistics,* a set of all the various values that individual observations may have and the frequency of their occurrence in the sample or population. **D7720**

3.1.6 *high nondetect, n*—a nondetect concentration with a highly elevated detection limit; for example, a concentration that resides in the upper decile of the analyte's distribution (that is, a detection limit above the 90th percentile of the data set). **E3382**

3.1.7 *median, n*—the 50th percentile in a population or sample. **E456**

3.1.8 *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. **D7048**

3.1.9 *outlier, n*—see *outlying observation*.

3.1.10 *outlying observation, n*—an extreme observation in either direction that appears to deviate markedly in value from other members of the sample in which it appears. **E178**

3.1.11 *parametric, adj*—a term referring to a statistical technique in which the distribution of the constituent in the population is assumed to be known. **D7048**

[3] 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.

3.1.12 *representative background concentrations, n*—chemical concentrations that are inclusive of naturally occurring sources and anthropogenic sources similar to those present at a sediment site but not related to current or historical site releases or activities. **E3382**

3.1.13 *representative background data set, n*—a background data set obtained by evaluating candidate background data using the steps described in Guide E3242. **E3382**

3.1.13.1 *Discussion*—The evaluation determines if there are any data points in the candidate background data set that are not representative of sediment site background conditions. These data points are then removed from the candidate background data set (using technically justifiable rationale) to obtain a representative background data set. Typically, this data set can be used to develop a BTV, which is a measure of the upper limit of representative background concentrations; it is this BTV that is often used as a representative background concentration.

3.1.14 *sediment(s), n*—a matrix of pore water and particles including gravel, sand, silt, clay and other natural and anthropogenic substances that have settled at the bottom of a tidal or nontidal body of water. **E3163**

3.1.15 *sediment site, n*—the area(s) defined by the likely physical distribution of COC(s) from a source area and the adjacent areas required to implement the corrective action. A site could be an entire water body or a defined portion of a water body. **E3240**

3.1.16 *upper tolerance limit (UTL), n*—the upper confidence limit (with specified confidence level) for a percentile of a distribution. **D7659**

3.1.16.1 *Discussion*—The UTL is the value below which a specified fraction of the population will be found, with a specified level of confidence. For example, the 95/95 UTL is a value for which one would have 95 % confidence that 95 % of the population is below the UTL.

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *arithmetic mean, n*—a measure of central tendency that is the sum of observed values in the sample divided by the sample size.

3.2.2 *background reference areas, n*—sediment areas that have similar physical, chemical, geological, biological, and land-use characteristics as the site being investigated but are not affected by current or historical site-related releases or activities.

3.2.3 *background threshold value (BTV), n*—a measure of the upper limit of representative background concentrations.

3.2.4 *cleanup level, n*—the prescribed average or point sediment concentration of a chemical that shall not be exceeded at the remediated site.

3.2.5 *conceptual site model (CSM), n*—the integrated representation of the physical and environmental context, the complete and potentially complete exposure pathways and the potential fate and transport of potential contaminants of concern at a site.

3.2.5.1 *Discussion*—The CSM should include both the current understanding of the site and an understanding of the potential future conditions and uses for the site. It provides a method to conduct the exposure pathway evaluation, inventory the exposure pathways evaluated, and determine the status of the exposure pathways as incomplete, potentially complete, or complete.

3.2.6 *false negative error, n—in statistical tests*, also known as "Type II" error.

3.2.6.1 *Discussion*—For the purposes of this guide, in site versus background comparisons, the error that occurs when the statistical procedure does not indicate concentrations above background, when such concentrations are present.

3.2.7 *false outlier, n*—measurements that are very large or small relative to the rest of the data but represent true extreme values of a distribution and indicate more variability in the population than was expected. (**3**)

3.2.8 *false positive error, n—in statistical tests*, also known as "Type I" error.

3.2.8.1 *Discussion*—For the purposes of this guide, in site versus background comparisons, the error that occurs when the statistical procedure indicates concentrations above background, when such concentrations are not present.

3.2.9 *population, n—as used in statistics*, a comprehensive set of values consisting of all possible observations or measurements of a certain phenomenon from which a sample is to be drawn.

3.2.10 *potential contaminant of concern (PCOC), n*—a contaminant whose sediment concentrations at the site may exceed applicable screening levels; this includes chemicals of potential environmental concern (COPECs) and chemicals of potential concern (COPCs).

3.2.11 *probability plot, n*—a plot of ascending observations in a sample, versus their corresponding cumulative probabilities, based on a specified distribution function.

3.2.12 *reference element, n*—a major element that represents the mineral to which a trace element may be adsorbed.

3.2.13 *sample, n—as used in statistics*, a group of observations taken from a population that serve to provide information that may be used as a basis for making a decision concerning the population.

3.2.14 *sample size, n—as used in statistics*, the number of observations or measurements in the sample.

3.2.15 *significance level, n—as used in statistical hypothesis testing*, the probability of rejecting a null hypothesis when it is true.

3.2.15.1 *Discussion*—Also known as "alpha" (α), it is selected prior to performing a statistical test. The significance level is commonly set to 0.05, but should be determined on a site-specific basis; consultation with a statistician to choose the optimal significance level may be warranted.

3.2.16 *tolerable error rate, n*—the specified maximum acceptable error rate set by the decision maker.

3.2.17 *trace element, n*—an element defined as generally being present at less than 0.1 weight percent in the sediment sample; its natural concentrations are typically one or more orders of magnitude lower than those of the reference elements.

3.2.18 *true outlier, n*—measurements that are very large or small relative to the rest of the data, but are a result of transcription errors, data-coding errors, or measurement system problems; or it is not representative of the investigated data population as confirmed by other lines of evidence.

3.2.19 *upper confidence limit (UCL), n*—an upper limit of an estimated value, such as the mean, which has a specified probability of including the true value, with a specified confidence level.

3.2.20 *upper percentile, n*—the value below which a specified percentage of observed values falls.

3.2.21 *upper prediction limit (UPL), n*—the value below which a specified number of future independent measurements will fall, with a specified confidence level.

4. Significance and Use

4.1 *Intended Use:*

4.1.1 This guide may be used by various parties involved in sediment corrective action programs, including regulatory agencies, project sponsors, environmental consultants, toxicologists, risk assessors, site remediation professionals, environmental contractors, and other stakeholders.

4.2 *Importance of the CSM:*

4.2.1 The CSM should be continuously updated and refined to describe the physical properties, chemical composition and occurrence, biologic features, and environmental conditions of the sediment corrective action project (Guide E1689).

4.3 *Reference Material:*

4.3.1 This guide should be used in conjunction with other ASTM guides listed in 2.1 (especially Guides E3344 and E3382); this guide should also be used in conjunction with the material in the References at the end of this guide (including **1**). Utilizing these reference materials will direct the user in developing representative background concentrations for a sediment site.

4.4 *Flexible Site-Specific Implementation:*

4.4.1 This guide provides a systematic, but flexible, framework to accommodate variations in approaches by regulatory agencies and by the user based on project objectives, site complexity, unique site features, regulatory requirements, newly developed guidance, newly published scientific research, changes in regulatory criteria, advances in scientific knowledge and technical capability, and unforeseen circumstances.

4.5 *Regulatory Frameworks:*

4.5.1 This guide is intended to be applicable to a broad range of local, state, tribal, federal, or international jurisdictions, each with its own unique regulatory framework. As such, this guide does not provide a detailed discussion of the requirements or guidance associated with any of these regulatory frameworks, nor is it intended to supplant applicable regulations and guidance. The user of this guide will need to be aware of the regulatory requirements and guidance in the jurisdiction where the work is being performed.

4.6 *Systematic Project Planning and Scoping Process:*

4.6.1 When applying this guide, the user should undertake a systematic project planning and scoping process to collect information to assist in making site-specific, user-defined decisions for a particular project, including assembling an experienced team of project professionals. These practitioners should have the appropriate expertise to scope, plan, and execute a sediment data acquisition and analysis program. This team may include, but is not limited to, project sponsors, environmental consultants, toxicologists, site remediation professionals, analytical chemists, geochemists, and statisticians.

4.7 *Use of Representative Background to Set a Boundary:*

4.7.1 Representative background concentrations for sediments can be used to delineate a sediment corrective action, establishing the boundary of the sediment corrective action area by distinguishing site-related impacts from representative background concentrations. This application requires the development of a BTV for the representative background data set.

4.8 *Use of Representative Background to Establish Cleanup Levels:*

4.8.1 Representative background concentrations for sediments can be used to establish cleanup levels for use in sediment corrective actions. In cases where risk-based sediment cleanup levels are below representative background concentrations, background concentrations are typically used as the cleanup level (**4**). This ensures that the cleanup levels are sustainable. Any recontamination from ongoing sources will eventually result in surface sediment concentrations greater than the risk-based cleanup level, but the surface sediment should still meet a cleanup level based on representative background concentrations, even following recontamination.

4.9 *Use of Representative Background in Risk Assessments:*

4.9.1 Representative background concentrations can be used in the risk assessment process (including human and ecological risk assessments) to understand risks posed by background levels of contaminants to human health and the environment, and the incremental risks posed by site-related releases or activities (or both) that result in sediment concentrations that exceed representative background concentrations. Conversely, they can be used to estimate the risk reduction for various contaminants, if sediment is remediated from existing COC concentrations to lower values (that is, representative background concentrations).

4.10 *Use of Representative Background in Post-Remedy Monitoring Programs:*

4.10.1 Post-remedy monitoring programs can also use representative background sediment concentrations either as a corrective action target or to understand how post-remedy concentrations compare to the sources not attributable to current or historical site releases or activities. Typically, source control actions taken to ensure that site-related releases are controlled and will not re-contaminate the post-corrective action sediments must be developed based on an understanding of ongoing contributions from representative background. Ongoing sources unrelated to current or historical site-related releases or activities (that may or may not be subject to source control actions) must be considered in this evaluation.

4.11 *Other Considerations:*

4.11.1 This guide does not cover all components of a program to develop representative sediment background concentrations.

4.11.2 The overarching process to develop representative background concentrations (including CSM considerations) is not covered in detail in this guide but is discussed in more depth in Guide E3382.

4.11.3 The selection of a background reference area(s) for the sediment site is not covered in detail in this guide but is extensively described in Guide E3344.

4.11.4 Sediment sampling and laboratory analyses are not covered in this guide. Guides E3163 and E3164 contain extensive information concerning sediment sampling and laboratory analyses.

4.11.5 Data quality objectives are not covered in this guide. Data quality objectives are described in (**5**).

4.11.6 Background study design considerations are not covered in this guide but are described in other references, including Guides E3163 and E3164, as well as (**6, 7**).

4.11.7 Geospatial analysis considerations are not thoroughly discussed in this guidance but are discussed in more depth relative to environmental evaluations in (**8**), which focuses on quality assurance concerns relative to geospatial analyses.

4.11.8 In this guide, only the concentrations of COCs are considered to be in scope. Residual background radioactivity is out of scope.

4.12 *Structure and Components of This Guide:*

4.12.1 The user of this guide should review the overall structure and components of this guide before proceeding with use, including:

Section 1	Scope
Section 2	Referenced Documents
Section 3	Terminology
Section 4	Significance and Use
Section 5	Overview of Representative Background Concentration Development Process
Section 6	Development of Candidate Background Data Sets
Section 7	Evaluation of Candidate Background Data Sets to Obtain Representative Background Data Sets
Section 8	Data Visualization
Section 9	Evaluation of High Nondetect Data Points
Section 10	Evaluation of Outlying Data Points
Section 11	Forensic Chemistry Evaluation of Organic Contaminants
Section 12	Geochemical Evaluation of Metals
Section 13	Methodology Application to Develop a Representative Background Data Set from a Candidate Background Data Set
Section 14	Development of Representative Background Concentrations
Section 15	Comparison of Sediment Site and Representative Background Data Sets Using Statistical Two-Sample Testing
Section 16	Keywords
Appendix X1	Organic and Inorganic Chemistry Overview
Appendix X2	Illustrative Case Studies from One Example Sediment Site
Appendix X3	Summaries for Outlier Testing and Two-Sample Statistical Testing
References	

5. Overview of Representative Background Concentration Development Process

5.1 *Importance of Representative Background:*

5.1.1 Multiple sources may contribute to the nature and extent of contamination at sediment sites. The largest contribution of contamination at sediment sites is typically attributed to current or historical site releases or activities. However, contamination can also result from natural or ongoing anthropogenic sources (or both) not related to current or historical site releases or activities. Discharges from combined sewer overflows (CSOs), industrial outfalls, and storm sewer systems (municipal and private) or surface runoff are examples of ongoing anthropogenic sources that may be unrelated to current or historical site releases or activities.

5.1.2 The off-site contamination not associated with current or historical site releases or activities is considered a component of representative background concentrations and will continue to be a source of contamination to the sediment site unless all transport pathways are eliminated. A primary objective of determining representative background concentrations is to account for any background chemical input (both natural and anthropogenic) that is expected to continue migrating onto the sediment site after the completion of corrective actions. One of the important principles for management of contaminated sediment sites is the control of sources of contamination, to the greatest extent practicable, prior to the initiation of corrective actions at the subject site (**4, 9**). However, it is rarely practicable to control all background sources.

5.1.3 Technically defensible representative background concentrations are those that accurately reflect chemical inputs to a sediment site from natural and ongoing anthropogenic sources unrelated to current or historical site releases or activities. In addition to informing or establishing technically defensible cleanup levels, representative background concentrations can assist in determining site boundaries, identifying COCs, establishing and optimizing realistic post-remedy monitoring plans, and assessing the performance of corrective actions.

5.1.4 In the absence of representative background concentrations, risk-based cleanup levels may be inappropriately used at sediment sites where representative background concentrations are actually greater than the risk-based cleanup levels. Similarly, if the representative background concentrations have been erroneously developed (for example, by the inappropriate exclusion of some outlier data points [false outliers]; refer to Section 10), inappropriately low cleanup levels could be used in the corrective action evaluation process. Under both circumstances, surface sediments at sediment sites will eventually return to representative background concentrations at some time after corrective actions are completed and cleanup levels will be exceeded. Due to exceedances of the inappropriately low cleanup levels, the corrective actions would be perceived as failures.

5.1.5 Attempting to implement corrective actions to achieve concentrations less than representative background is not sustainable over the long-term and can require considerable expenditures that serve no environmental or public health purpose (Guide E3382). The process described in Guide E3382 is intended to help promote a scientifically sound approach for developing representative background concentrations, leading to corrective action decisions that avoid costly perceived corrective action failures at sediment sites. The topics covered in this guide are a critical component of the process outlined in Guide E3382 and include developing candidate background data sets; data visualization of candidate background data sets; evaluation of candidate background data sets to develop representative background data sets using statistical, forensic chemistry and geochemical methodologies; the development of various measures of representative background concentrations for applications at sediment sites using representative background data sets; and the application of two-sample statistical tests to compare representative background and sediment site data sets.

5.2 *Overview of Process to Develop Representative Background Concentrations in Sediment:*

5.2.1 Application of background guidance for soil and groundwater at upland sites may not be appropriate at sediment sites. Sediment sites have many different characteristics that are not present at upland sites (Guide E3248), including physical characteristics, geochemical characteristics, biological characteristics, and different contaminant emplacement and transport mechanisms.

5.2.2 This guide and its associated guides (Guides E3163, E3164, E3240, E3344, and E3382) have been developed (in part) to fill a gap due to the absence of existing guidance from various regulatory agencies for the development of representative background concentrations for contaminated sediment sites.

5.2.3 Fig. 1 presents the overall framework to develop the BTV, which is a measure of the upper limit of representative background concentrations at a sediment site; this process is presented in detail in Guide E3382. As a first step, a thorough understanding of the sediment site is necessary before developing the BTV. This can be accomplished by developing a sediment site CSM (refer to Guides E3240 and E3382). As part of this CSM, the sediment site PCOCs must be identified.

5.2.4 Once the preliminary sediment site CSM has been developed, a suitable background reference area (or areas) can be identified for sampling (that is, the second step in Fig. 1); the methodology used to do this is presented in detail in Guide E3344.

5.2.5 Candidate background data sets are typically obtained in two ways: *(1)* collecting sediment samples from background reference areas that have characteristics as similar as possible to that of the sediment site (see Guide E3344 for a detailed discussion on the selection of the background reference area), or *(2)* extracting candidate background data sets from the sediment site data from portions of the site that have been unaffected by current or historical site releases or activities (see Appendix X2 and Ref. (**7**) for a detailed discussion of background data extraction from the sediment site data set). Additionally, under certain circumstances data sets from *(1)* and *(2)* can be combined to develop a single candidate background data set. Section 6 describes the collection of sediment samples from background reference areas and extraction of background data sets from the sediment site data to develop candidate background data sets.

5.2.6 Once a candidate background data set is developed, Sections 6 – 14 describe the process used to evaluate these data sets to develop representative background data sets for the site and then develop BTVs for these representative background data sets (see Section 7 for further details on the evaluation process).

6. Development of Candidate Background Data Sets

6.1 *Background Reference Area Characteristics:*

6.1.1 The background reference area(s) should have similar physical, chemical, geological, biological, and land-use characteristics as the sediment site. Additionally, such areas should not be influenced by current or historical site-related releases or activities but should include ongoing sources similar to those present at the sediment site, as well as a similar land use (Guide E3344). For example, if the sediment site is located in an industrial area with CSOs, the background reference area(s) should be located in an industrial area with CSOs. The more developed the CSM, the more informed the choice of background reference areas will be. Additional information on the selection process is provided in Guide E3344 and (**10**).

6.1.2 Once an appropriate off-site background reference area(s) is identified for the sediment site, existing sediment sample data (if available) should be acquired for the background reference area(s) and evaluated for data quality and usability. Existing data from a previous study may be suitable for inclusion in the candidate background data set or to inform the study design for a new background reference area sediment sample collection program. However, the inclusion of existing data in the candidate background data set must be assessed on a case-by-case basis (**7**).

6.1.2.1 Inclusion of existing data in the candidate background data set is inappropriate if different analytical test methods were used to generate the existing candidate background data set and the sediment site data set. Background data sets generated by different analytical laboratories should also be subjected to a careful assessment of laboratory standard operating procedures to ensure comparability of results. Additional statistical comparison of results generated from the different laboratories are also recommended. The objective is to have comparable background data, even if the data are obtained from different laboratories.

6.2 *Collection of Candidate Background Data Sets from Background Reference Areas:*

6.2.1 Data collection can commence once the background reference area(s) are identified. This guide does not address methods and means of data collection. However, sample collection and laboratory analysis methods at the background reference area(s) should be as similar as possible to those used during sediment site data collection. When collecting data from a background reference area(s) for comparison to sediment site data, it is critical for background data to be collected in the same manner (such as, from the same depth interval) as the sediment site under investigation and to use the same analytical test methodology that was used to determine compliance with cleanup levels (ideally, using the same laboratory that analyzed

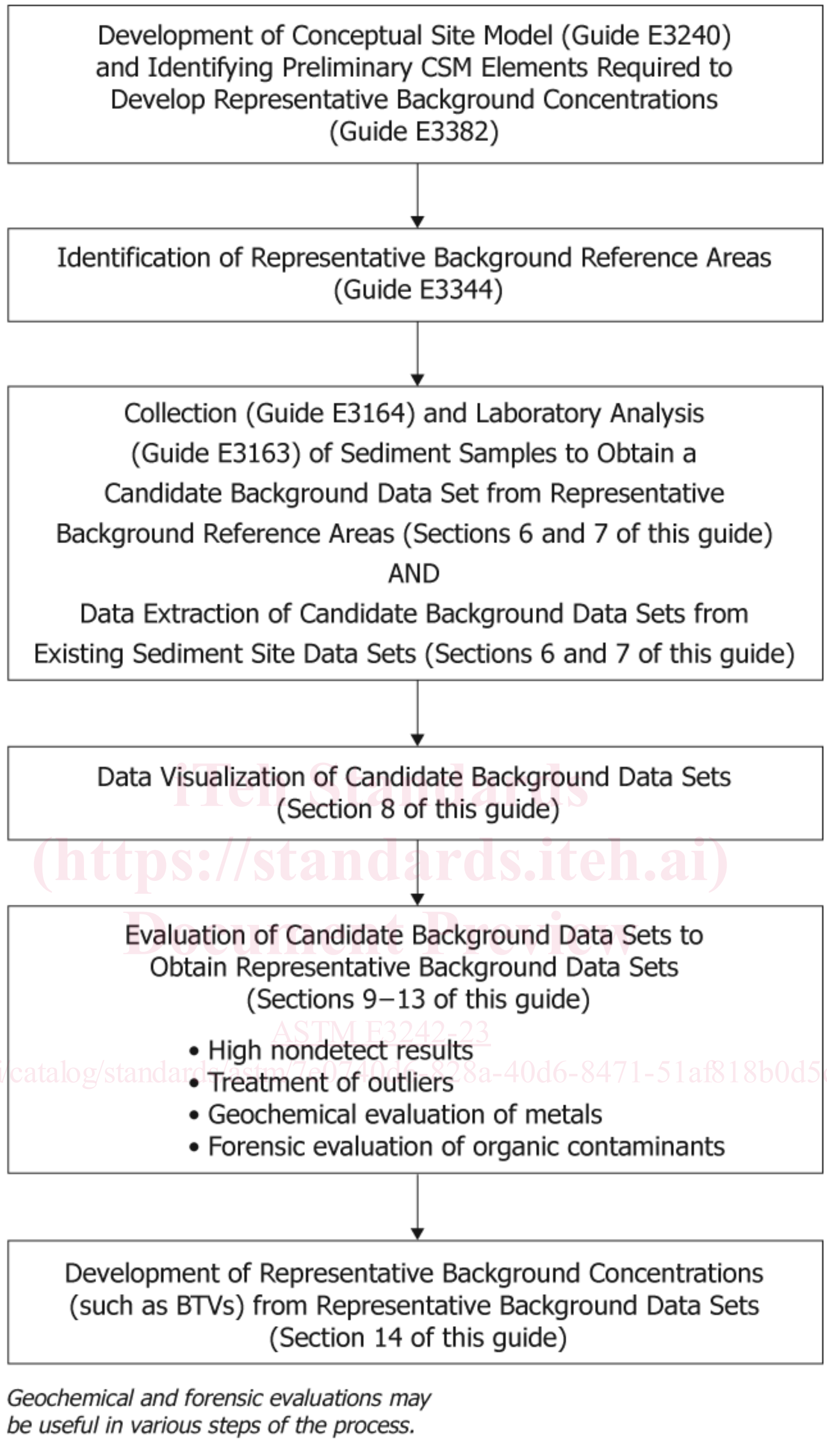

FIG. 1 Process to Develop Representative Sediment Background Concentrations (Modified from Guide E3382)

the samples from the sediment site). This allows a direct comparison between chemical concentrations in background reference area samples to sediment site samples—a crucial step in the development of representative background concentrations.

6.2.2 The optimal number of background reference area sediment samples needs to be determined on a project-specific basis by qualified personnel on the project team.

6.2.3 New sediment samples from the off-site background reference area(s) should be collected and sent to the laboratory for analysis of PCOCs and other parameters (such as total organic carbon and grain size distribution). Upon validation, the laboratory results and suitable existing data would constitute the candidate background reference area data set. Detailed sediment sampling and analysis guidance is provided in Guides E3163 and E3164; guidance on various sampling designs is provided in (**6**).

6.2.4 Regulatory agency agreement on the scope and scale of the background reference area sampling effort to obtain a candidate background data set is important and should be captured in a site's Data Quality Objectives. In the United States, the use of the USEPA's Data Quality Assessment approach (**3**) is recommended if no superseding regulatory guidance is available.

6.2.5 Concentrations from background reference areas may be characteristic of one or more statistical populations with distinct features. For example, sediment background concentrations from a basin surrounded by urban developments or industrial areas (or both) will be distinctly different from those collected from another portion of the same basin surrounded by agricultural areas. Combining background data sets that represent different statistical populations can lead to erroneous or misleading results. Candidate background data sets are those that are collected from sampling locations with physical, chemical, geological, biological, and land-use characteristics most similar to the sediment site.

6.3 *Extraction of Candidate Background Data Sets from Sediment Site Data Sets:*

6.3.1 Although collecting samples from off-site background reference areas is typically preferable, in many instances (especially in urban areas), identification of such areas are problematic. Under such conditions, candidate background data sets may potentially be extracted from sediment site data, as long as part of the sediment site has not been impacted by current or historical site releases or activities.

6.3.1.1 Extracting candidate background data from sediment site data not only maximizes the utility of existing data, but also avoids the often complex task of selecting separate background reference areas that adequately display physical, chemical, geological, biological, and land-use characteristics similar to the site, as described in Guide E3344.

6.3.2 Extraction of candidate background data from sediment site data often involves utilizing probability plots to segregate site data into impacted versus unimpacted populations for each COC (see Appendix X2).

6.3.2.1 Even when data from separate off-site background reference areas are available, an extracted site-specific background data set can provide (if certain conditions are met) additional data for inclusion in the candidate background data set. Therefore, an analysis of existing sediment site data is always recommended. A more complete review of this topic is presented in (**7, 11**).

7. Evaluation of Candidate Background Data Sets to Obtain Representative Background Data Sets

7.1 *Overview:*

7.1.1 Candidate background data sets for the sediment site are generated by sampling a background reference area(s) or using candidate background data extracted from the sediment site data set; or a combination of both if certain conditions are met (see Appendix X2).

7.2 *Evaluation Methodologies Summary:*

7.2.1 Once developed, candidate background data sets are then evaluated using the methodologies outlined in the following sections:

7.2.1.1 Visualization of the candidate background data set (Section 8).

7.2.1.2 Evaluation of high nondetect results (Section 9).

7.2.1.3 Evaluation of statistical outlying data points (Section 10).

7.2.1.4 Forensic chemistry evaluation of organic contaminants (Section 11).

7.2.1.5 Geochemical evaluation of metals (Section 12).

7.2.2 If technically justifiable, some data points may be excluded from the candidate background data set in the development of a representative background data set (Section 13).

7.3 *Uses of Representative Background Data Sets:*

7.3.1 A representative background data set can be used to develop a BTV for this data set (that is, a measure of the upper limit of representative background concentrations for the COC) as outlined in Section 14.

7.3.2 A representative background data set can also be used in two-sample statistical comparisons with the sediment site data set as described in Section 15.

8. Data Visualization

8.1 *Overview:*

8.1.1 Evaluation of a candidate background data set should begin by visualizing the data. There are a number of techniques described in this section that can be used, depending on the characteristics of the candidate background data set. These plots are exploratory in nature and not all plots shown are required when visualizing the candidate background data set. Different plots depicting the same data set are illustrated in the following subsections. Other tools (for example, geographic information system [GIS] post plots) can also be used for data visualization.

8.2 *Dot Plot:*

8.2.1 Dot plots represent each concentration in a candidate background data set as an individual dot (see Fig. 2), with concentration values listed along the *x*-axis (**12**). Samples with similar concentrations appear as vertical stacks, and large data sets can be accommodated by using one dot to represent a predetermined number of data points. This plot allows all data points to be viewed, and no distributional assumptions are imposed on the data. Symmetry, bimodal or multi-modal groupings, and skewness can be discerned in dot plots.

8.3 *Histograms:*

8.3.1 Histograms depict data sets in bar form, with concentrations (grouped in "bins," or intervals) along the *x*-axis and the corresponding frequency (or percentages of frequency) for each concentration interval along the *y*-axis (see Fig. 3). The area of each bar reflects the proportion of that concentration interval within the candidate background data set. Histograms

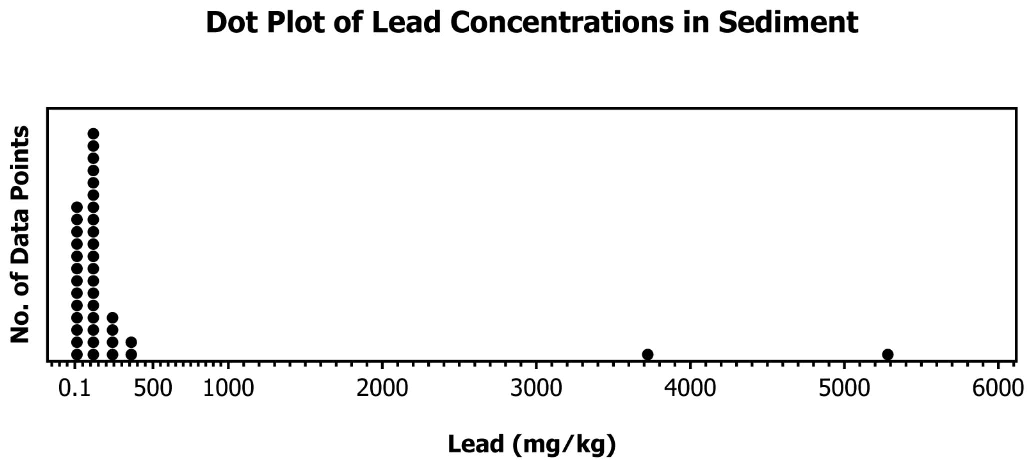

FIG. 2 Dot Plot of Lead Concentrations in a Candidate Background Data Set

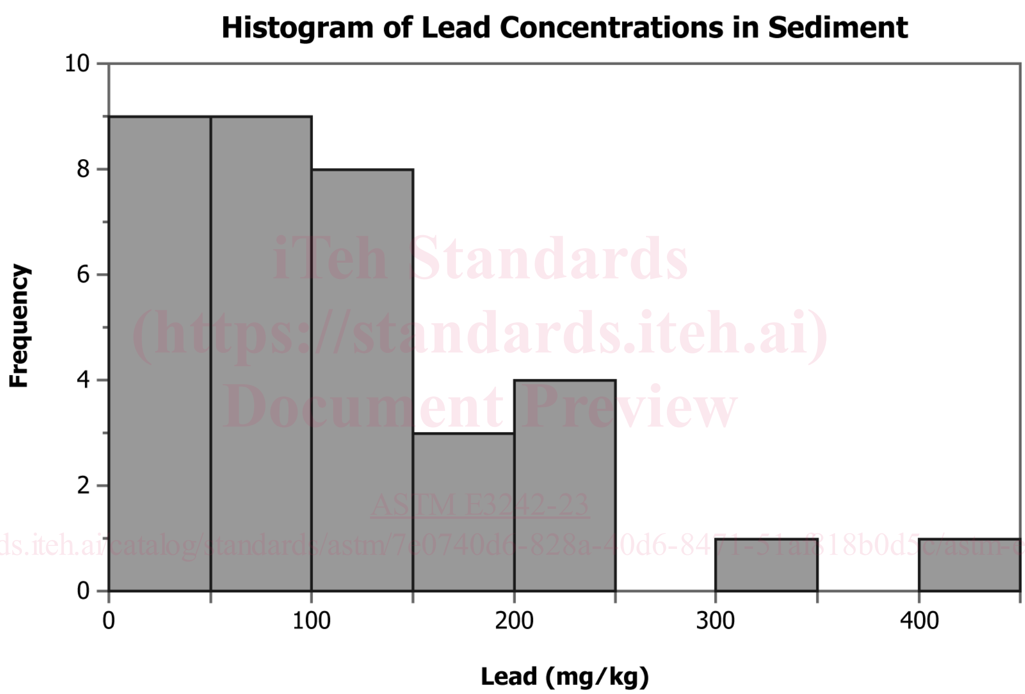

FIG. 3 Histogram of Lead Concentrations in a Candidate Background Data Set

are commonly used in conjunction with goodness-of-fit tests to assess the shape of a data distribution, because the graphs can reveal such features as symmetry, skewness, and bimodality. However, the observed shape of a data set is affected by the bin size (for example, 10 mg/kg interval versus 20 mg/kg interval), which should be carefully selected. Multiple data sets can be depicted on the same histogram using unique colors or patterns.

8.4 *Box Plots:*

8.4.1 Box plots, or box-and-whisker plots, are used to compare two or more groups of data (**3, 13**). A box plot (see Fig. 4) provides a summary view of an entire data set, including the range of concentrations, degree of symmetry, and skewness of the data. The box encloses the central 50 % of the data points ("interquartile range"), with the top of the box representing the 75th percentile and the bottom of the box representing the 25th percentile; the median is represented by a symbol within the box. In this example, the upper whisker extends to the maximum data point and the lower whisker extends to the minimum data point. A side-by side configuration of box plots permits visual comparison of multiple data sets to quickly discern whether the data distributions are similar or distinct.

8.4.2 Users can define the appearance of box plots. Default settings in statistical software programs typically identify predefined "outlier values" (for example, values outside 1.5 times the interquartile range) and "extreme values" (for example, values outside 3 times the interquartile range). These settings are arbitrary and can be avoided by simply extending the upper and lower whiskers to the maximum and minimum data points. See Section 10 for discussions of statistical outliers and their appropriate treatment.

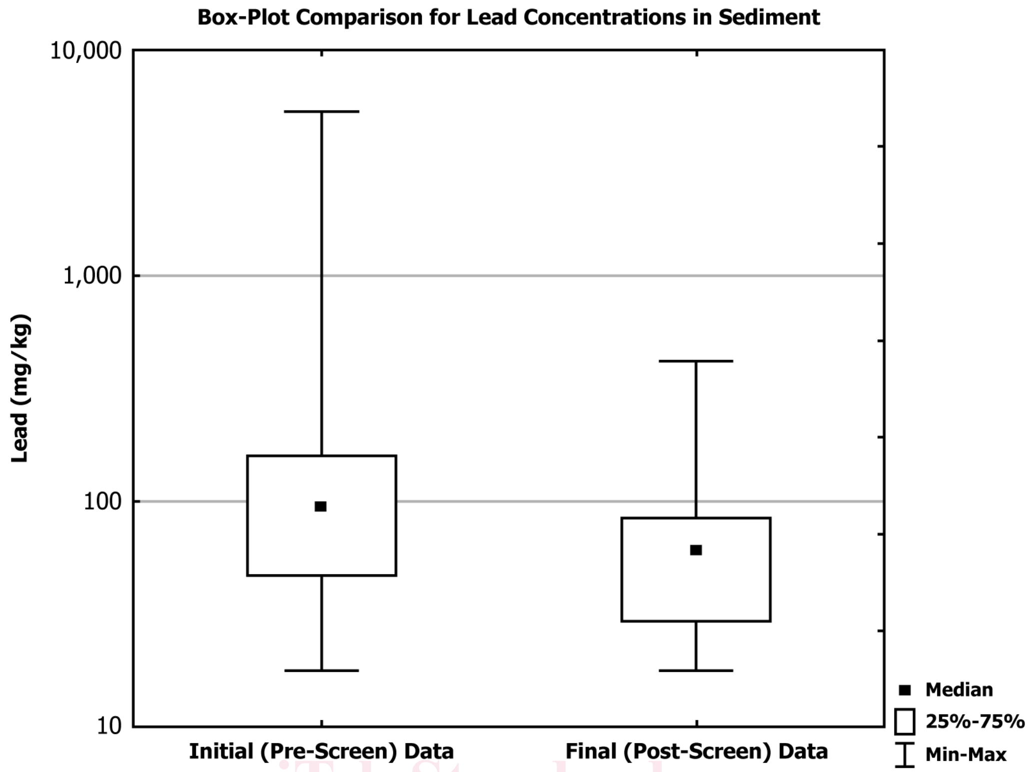

FIG. 4 Box Plots of Lead Concentrations in a Candidate Background Data Set (Left Side) and Following Evaluation in the Representative Background Data Set (Right Side)

8.5 *Probability Plots:*

8.5.1 A probability plot is often used to visualize data distributions and determine whether the data set is distributed consistently with a known distribution (such as, normally or lognormally distributed data) (**3**). Probability plots can indicate the presence of possible outlier values, and they can help determine whether the data set is derived from a single or multiple subpopulations (see Fig. 5). For example, a typical normal probability plot is constructed by plotting each observed data in an ascending order on the *y*-axis versus its corresponding standard normal probability values for its respective rank on the *x*-axis. A logarithmic scale on the *x*-axis is

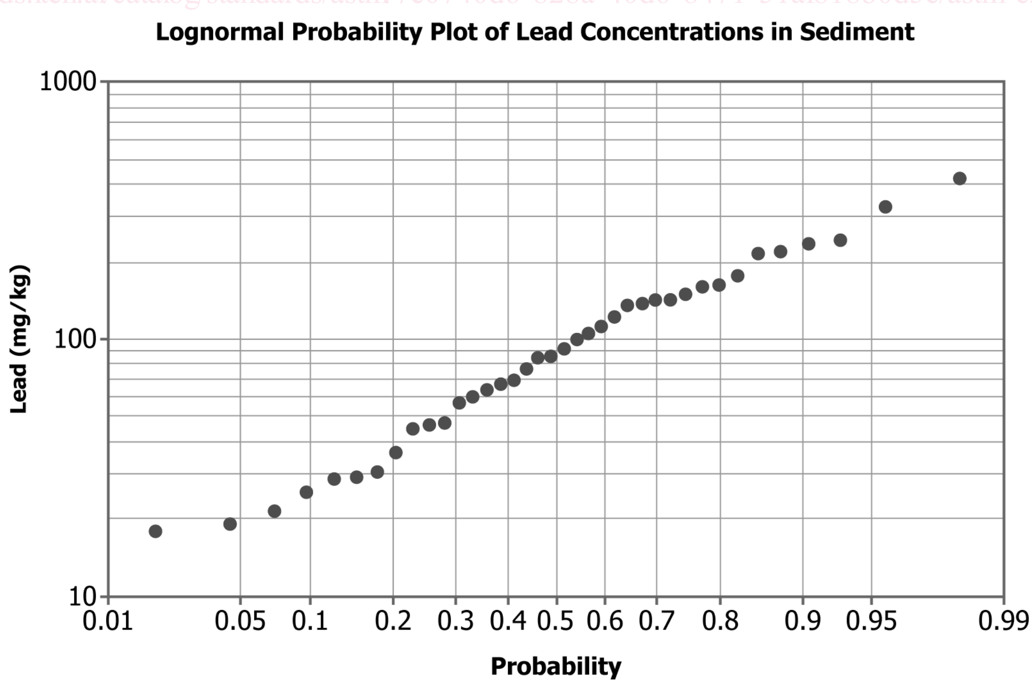

FIG. 5 Lognormal Probability Plot of Lead Concentrations in a Candidate Background Data Set

used for lognormal distributions (see Fig. 5), and normal scale is used on the *x*-axis for normal distributions. If the points fall roughly on a straight line, it can be concluded that the underlying data distribution is approximately normal (or lognormal if a logarithmic scale is used for the *x*-axis), with the slope proportional to the variance of the data. Many environmental data sets do not fit parametric models (see 14.1), such as the normal or lognormal distribution, so probability plots can have limited utility or may lead to erroneous conclusions for those data sets.

8.6 *Percentile Plots:*

8.6.1 Percentile plots are similar to probability plots but depict concentration versus percentile rather than concentration versus probability (see Fig. 6). The data are first rank-ordered and then concentrations are plotted on the *y*-axis with the corresponding percentiles plotted on the *x*-axis (**14**). Normally distributed data appear as a straight line if a linear concentration scale is used for the *y*-axis, and lognormally distributed data appear as a straight line if a logarithmic concentration scale is used; statistical outliers will appear above or below the linear trend. A break in slope may be observed if the distribution is bimodal or if multiple samples have identical concentrations. As with probability plots, percentile plots permit a qualitative assessment of the shape of the data. A key advantage of percentile plots is that they are nonparametric, so they can be used to visualize any data set without making assumptions regarding the distributional shape of the data set.

8.7 *Scatter Plots:*

8.7.1 In support of the geochemical evaluation of metals data, scatter plots are constructed to explore elemental associations and identify potentially contaminated samples (**15**). Trace element concentrations are plotted along the *y*-axis and the corresponding reference element concentrations (typically major elements such as iron and aluminum) are plotted along the *x*-axis (see Fig. 7). The reference element represents the mineral to which the trace element may be adsorbed, as discussed in 12.5. A common trend (not necessarily linear) is observed in the absence of contamination, due to similar trace-versus-reference element ratios. A sample with excess trace element from a contaminant source will exhibit an anomalously high trace-versus-reference element ratio relative to unimpacted samples and will lie above the trend.

8.8 *Ratio Plots:*

8.8.1 Ratio plots are recommended to accompany each scatter plot (**16**) in geochemical evaluations. Ratio plots depict trace element concentrations along the *y*-axis and the corresponding elemental ratios (that is, the trace element concentration divided by the reference element concentration for each sample) along the *x*-axis (see Fig. 8). Unimpacted samples exhibit consistent trace-versus-reference element ratios. A sample with excess trace element from a contaminant source will exhibit an anomalously high trace-versus-reference element ratio relative to unimpacted samples and will lie to the right of the unimpacted samples in the ratio plot.

9. Evaluation of High Nondetect Data Points

9.1 *Overview:*

9.1.1 Site and candidate background data sets often include samples with nondetect results. Nondetects occur in environmental data sets because laboratory methods used to measure contaminants are limited in their sensitivity.

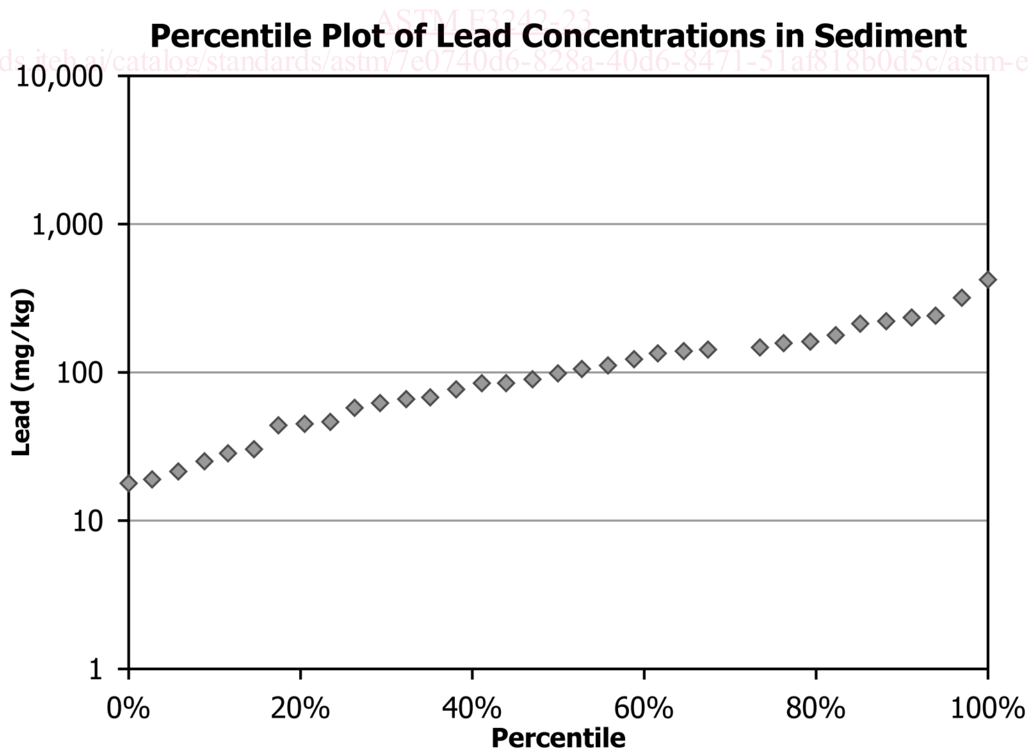

FIG. 6 Nonparametric Percentile Plot of Lead Concentrations in a Candidate Background Data Set

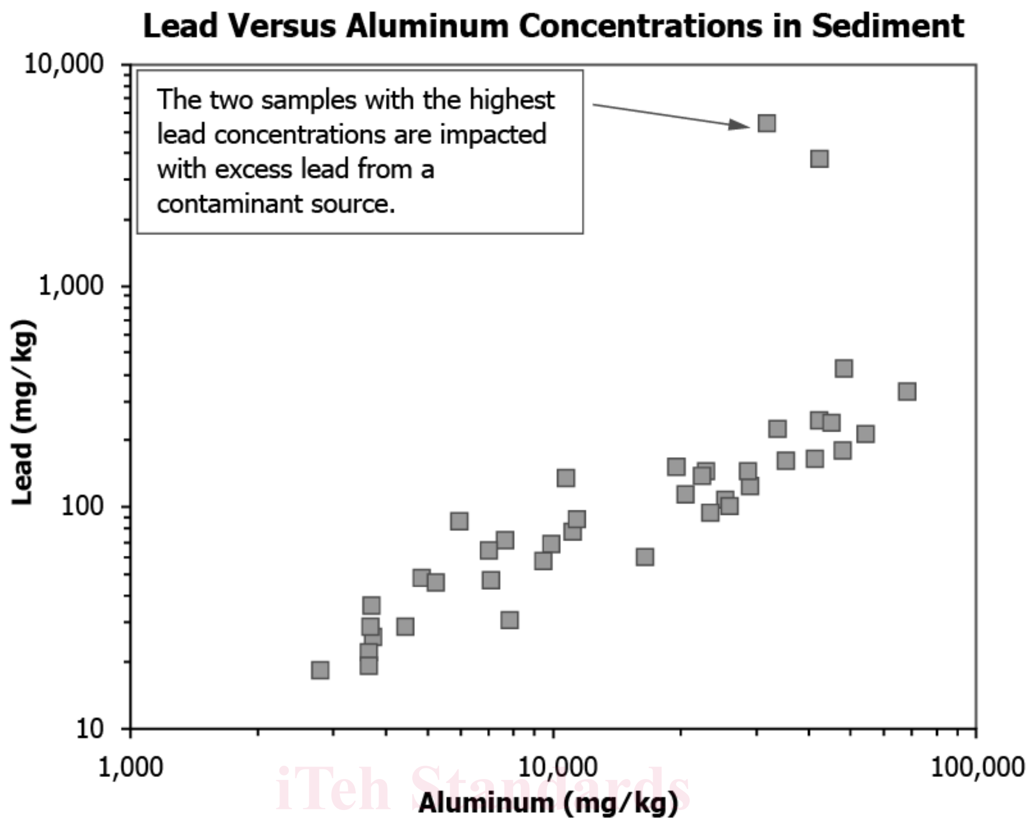

FIG. 7 Scatter Plot of Lead Concentrations Versus Aluminum Concentrations in a Candidate Background Data Set

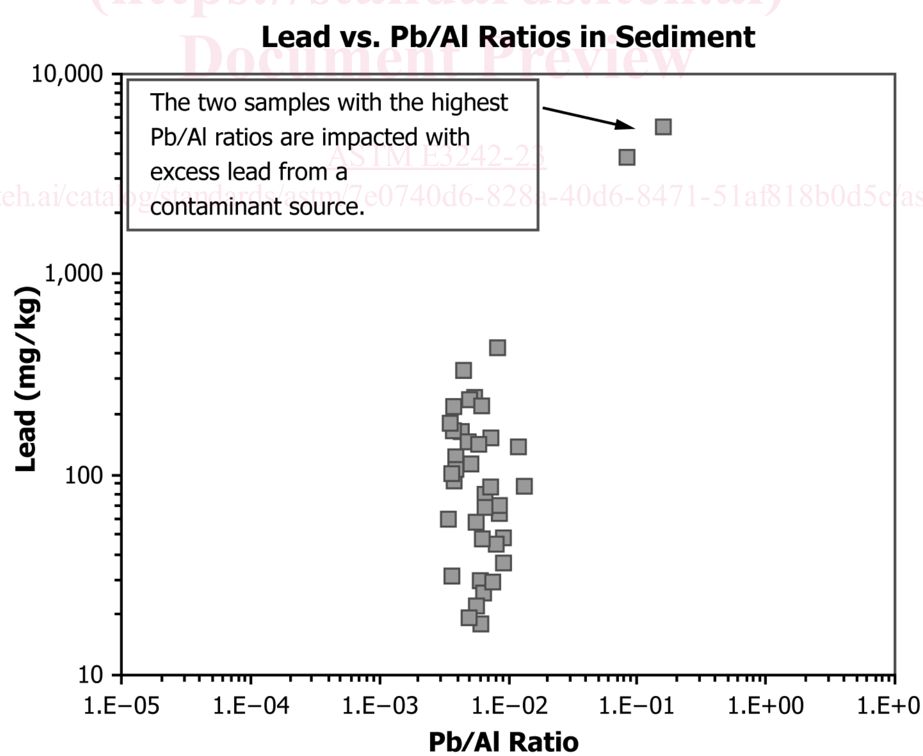

FIG. 8 Ratio Plot Depicting Lead Concentrations Versus Lead/Aluminum Ratios in a Candidate Background Data Set