Government Gazette

of the State of

New South Wales

Number 372–Environment
Friday, 12 August 2022

The New South Wales Government Gazette is the permanent public record of official NSW Government notices. It also contains local council, non-government and other notices.

Each notice in the Government Gazette has a unique reference number that appears in parentheses at the end of the notice and can be used as a reference for that notice (for example, (n2019-14)).

The Gazette is compiled by the Parliamentary Counsel's Office and published on the NSW legislation website (www.legislation.nsw.gov.au) under the authority of the NSW Government. The website contains a permanent archive of past Gazettes.

To submit a notice for gazettal, see the Gazette page.

Declaration of significantly contaminated land

Section 11 of the *Contaminated Land Management Act 1997*

Declaration Number 20221105; Area Number 3515

The Environment Protection Authority (EPA) declares the following land to be significantly contaminated land under s 11 of the *Contaminated Land Management Act 1997* (Act).

1. Land to which this Declaration applies

This declaration applies to significantly contaminated land described as Tarago former Station Masters Cottage, 106 Goulburn Street, (Lot 1 DP816626) Tarago, NSW 2580 within the local government area of Goulburn-Mulwaree Council (the Land).

The land to which this declaration applies is shown on the attached map and is shaded blue.

2. Substances affecting the Land

The EPA has reason to believe that the Land is contaminated with the following substances (Substances) in such a way as to warrant regulation as significantly contaminated land under the Act:

- Lead

3. Nature of harm caused by the Substances

The EPA has considered the matters in s 12 of the Act before making this declaration. The EPA has reason to believe harm has been, or may be, caused by the Substances, including:

- Lead concentrations in soil within the historic Station Masters Cottage (Lot 1 DP816626) exceed national guideline values for the protection of human health and the environment.

- There are potentially complete exposure pathways for onsite and offsite ecological receptors.

- Based on the current levels of contamination identified, the site is not appropriate for the existing land-use and remediation or management is required. Remediation will be required to facilitate residential land-use which it is zoned to do so under the Goulburn-Mulwaree Council LEP (2009).

- Lead levels in soil and dust were identified within the historic Station Masters Cottage at levels greater than the relevant assessment criteria.

- Lead, arsenic, chromium, copper, nickel and zinc were found on the rail corridor at concentrations exceeding national guidelines values for the protection of human health and environment which may have migrated to the Station Masters Cottage and as such should be assessed for.

4. Further action to carry out voluntary management under the Act

The making of this declaration does not prevent the carrying out of voluntary management of the Land by any person. Any person may submit a voluntary management proposal for the Land to the EPA.

5. Submissions invited

Any person may make a written submission to the EPA on:

- whether the EPA should issue a management order in relation to the Land; or
- any other matter concerning the Land.

Submissions should be made in writing and sent to:

Manager Regulatory Operations Regional South
Environment Protection Authority
Locked Bag 5022
PARRAMATTA NSW 2124

or emailed to info@epa.nsw.gov.au care of David Langston, by not later than 9 September 2022.

CATE WOODS
Director Regulatory Operations Regional
Environment Protection Authority

(By delegation)

Date: 3 August 2022

Further Information about this Declaration

Management Order may follow
If management of the Land or part of the Land is required, the EPA may issue a Management Order under s 14 of the Act.

Amendment or Repeal of Declaration
This declaration may be amended or repealed. It remains in force until it is otherwise amended or repealed. The subsequent declaration must state the reasons for the amendment or repeal (s 44 of the Act).

Information recorded by the EPA
Section 58 of the Act requires the EPA to maintain a public record. A copy of this significantly contaminated land declaration will be included in the public record and is available for access at the principal office of the EPA and on the EPA's website.

Information recorded by Councils
Section 59(a) of the Act requires the EPA to inform the relevant local Council as soon as practicable of this declaration. Pursuant to s 59(2)(a) of the Act, land being declared to be significantly contaminated land is a prescribed matter to be specified in a planning certificate issued pursuant to s 10.7 of the *Environmental Planning and Assessment Act 1979*. The EPA is also required to inform the relevant Council as soon as practicable when the declaration is no longer in force. Pursuant to s 59(3) of the *Contaminated Land Management Act 1997*, if a Council includes advice in a planning certificate regarding a declaration of significantly contaminated land that is no longer in force, the Council is to make it clear on the planning certificate that the declaration no longer applies.

Relationship to other regulatory instruments

This declaration does not affect the provisions of any relevant environmental planning instruments which apply to the land or provisions of any other environmental protection legislation administered by the EPA.

Image: Area of proposed declaration is coloured and highlighted blue.

The original image was taken from https://maps.six.nsw.gov.au/ on 2 May 2022 and adapted by the NSW EPA.

CONTAMINATED LAND MANAGEMENT ACT 1997

Notice of Guidelines

'Sampling Design: Contaminated Land Guidelines'

I hereby give notice under section 105(1) of the *Contaminated Land Management Act 1997*, that the 'Sampling Design: Contaminated Land Guidelines' are made. This guideline takes effect under section 105(2)(c) of the *Contaminated Land Management Act 1997* upon publication in the *Government Gazette*.

The 'Sampling Design: Contaminated Land Guidelines' revokes the 'Sampling Design Guidelines: Contaminated Sites September 1995'.

27 July 2022

Clarence Brown
Acting Director Regulatory Policy and Reform
Environment Protection Authority

Environment Protection Authority

Sampling design part 1 - application

Contaminated Land Guidelines

Acknowledgements

The NSW Environment Protection Authority acknowledges the contributions of the following organisations in the preparation of these guidelines:

Easterly Point Environmental Pty Ltd, JBS&G Australia Pty Ltd, Enviroview Pty Ltd, GHD Pty Ltd, Envirocene Pty Ltd, WSP Australia Pty Ltd, JK Environments Pty Ltd.

Contents

Figures

Tables

These guidelines are intended to assist with the design, review or regulation of sampling programs for contaminated sites.

The guidelines help identify and mitigate risks to human health and the environment in the design of appropriate sampling and analysis plans.

1. Introduction

The NSW Environment Protection Authority (EPA) has prepared these guidelines to assist contaminated land consultants, site auditors, regulators, planning authorities, landholders and developers, and inform members of the public who have an interest in the outcomes of the assessment and management of contaminated land. They will help consultants to design sampling programs for contaminated sites, regarding where samples are collected, how many samples are collected, and ways the data are compared to relevant criteria.

The guidelines will help users obtain data that is appropriately representative for the purposes of the sampling and the media being sampled, and carry out the subsequent analysis and interpretation of the collected data.

As when following any guidance, users should justify the approaches they use, and demonstrate that they are appropriate and fit for purpose.

These guidelines replace the EPA's previous sampling design guidelines: Environment Protection Authority (EPA) 1995, *Contaminated sites: sampling design guidelines*, EPA 95/59, NSW EPA, Sydney.

The guidelines are in two parts. The first part (this document) describes the application of sampling design; the second part provides guidance on interpreting the results.

1.1. Background

Assessment of site contamination is risk based and should take a weight of evidence approach. A major objective of contamination assessment is to determine the nature and extent of contamination by collecting representative environmental samples for characterisation and chemical analysis. The type of sampling carried out, and the methods used to analyse and interpret the resulting data, significantly influence assessment's validity.

This document provides specific recommendations and procedures for consultants and reviewers of site investigations. However, it is not all-encompassing. For methods it does not describe, or for more complex problems, refer to other relevant guidelines and information sources – this document points to many – or consult an environmental statistician.

These guidelines should be used at the beginning of the site investigation, when the preliminary conceptual site model (CSM) has been developed and data gaps for site characterisation have been identified. The next steps are to establish the data quality objectives (DQOs), then identify the processes that could have resulted in contamination, the contaminants of potential concern (CoPC) and the target media for the investigation.

1.2. Scope of these guidelines

Section 2

Introduction to systematic planning, including CSMs and DQOs. Additional information on DQOs is provided in Appendix A: DQOs and the environmental data life-cycle process, and a hypothetical worked example is given in Appendix B: Data-quality objectives: worked example.

Section 3

General considerations regarding environmental sampling and statistical aspects in site contamination assessment.

Section 4

Objectives of sampling programs, including a discussion of ways characterisation and validation correspond, modes of contamination and sampling objectives.

Section 5

Main sampling strategies and considerations for soil and fill, both in situ and in stockpiles, and brief discussion of other environmental media such as groundwater, surface water, sediments and ground gases, including references to more detailed guidance.

Section 6

Methods for detecting significantly elevated concentrations of contamination (that is, hotspots). Appendix C: Determining sampling grids for hotspot detection provides methods for determining sampling grids for hotspot detection, including the recommended grid sizes for characterisation using a systematic sampling pattern.

Section 7

The number of samples required, including existing guidance and statistical tests for determining the number of samples, using the combined risk value (CRV) method and the maximum probable error (MPE) method. Appendix D: Summary of existing guidance for sample design summarises guidance regarding sampling design. Procedures and worked examples using the CRV and MPE methods are provided in Appendix E: Determining the number of samples by the CRV method and Appendix F: Determining the number of samples by the MPE method.

1.3. Legal framework, policy and relationship to other guidelines

These guidelines are made under the *Contaminated Land Management Act 1997* (CLM Act). They should be read in conjunction with the CLM Act, the Contaminated Land Management Regulation 2013 (CLM Regulation), and any guidelines made or approved by the EPA under the CLM Act.

These guidelines complement other guidelines made by the EPA, and several national guidance documents that have been approved by the EPA. Those guideline documents are listed in the reference section and are specifically referenced in the text, where appropriate.

Sampling of materials must comply with any other relevant statutory requirements that apply, including those set out in statutory instruments. For example, materials intended to be used offsite may need to comply with requirements set out in Resource Recovery Orders and Exemptions. For more information about orders and exemptions, please refer to the EPA website.

1.4. Environmental media

These guidelines address soil and solid media sampling, as these are the most common targets when assessing site contamination. They also provide limited information about other media, including groundwater, surface water, sediments and air. Some of the statistical procedures described in these guidelines can be applied to these media, however the EPA recommends that the following references are consulted when designing sampling programs for other media:

- NEPC 2013 – soil, groundwater and soil vapour
- ANZG 2018 – surface water
- DEC 2007 – groundwater
- EPA 2020a – ground gases
- DECCW 2010 – soil vapour
- CRC Care 2013 – soil vapour

- Simpson and Batley 2016 – sediments.

General advice is provided on sampling for emerging contaminants, along with specific references.

This document does not provide detailed advice on sampling methods or techniques for collecting samples.

This document does not specifically address biota sampling and ecotoxicity testing. For these specialty areas, see the following references:

- Australian and New Zealand Governments and Australian State and Territory Governments (ANZG) 2018, *Australian and New Zealand guidelines for fresh and marine water quality*, Water Quality Australia, Canberra ACT. Available at: www.waterquality.gov.au/anz-guidelines.
- Department of Environment and Science (DES) 2018, *Monitoring and sampling manual: environmental protection (water) policy 2009*, Queensland Department of Environment and Science, Brisbane.
- Department of Environment and Conservation (DEC) 2004*, Australian river assessment system (AUSRIVAS) sampling and processing manual 2004*, NSW Department of Environment and Conservation, Sydney.

1.5. Decision makers

Various decision-makers are concerned with the outcomes of site contamination assessment projects and investigation results, including:

- the consultant team
- clients, landowners, property developers
- accredited site auditors
- planners and other technical specialists
- regulators, including the local government, EPA and other state government bodies
- other relevant stakeholders, such as adjacent landholders, the local community, and non-government organisations.

When planning and reporting on site contamination assessment projects, consultants should recognise that other decision makers are not always technical specialists. Therefore, the methods used in collecting and analysing site contamination assessment sampling data should be clearly documented and discussed.

Other decision makers may impose limitations on the extent of investigations, such as costs and access restrictions set by clients, or heritage and ecology restrictions imposed by regulators. In all instances, clear and appropriate explanation and justification of the implemented sampling program should be provided, including the benefits of the approach selected, along with relevant assumptions, limitations and remaining data gaps.

Importantly, where the sampling program deviates from made or approved guidance, this should be clearly articulated, including the rationale and justification for any such deviations.

2. Systematic planning

A systematic planning process should be used to define the objectives of all site assessment, remediation and validation programs, and to develop sampling and analysis quality plans (SAQPs) for collecting and evaluating representative data to achieve those objectives. The *National Environment Protection (Assessment of Site Contamination) Measure* (NEPC 2013, B2) recommends the use of conceptual site models (CSMs) and DQOs for systematic planning. CSMs should be prepared in accordance with EPA 2020b.

2.1. Conceptual site models

CSMs provide a spatial and temporal overview of the contamination at sites and their surroundings. They highlight the contaminant sources and potential receptors, and the potential exposure pathways between the sources and receptors. Robust CSMs should include the known and potential:

- sources of contamination and contaminants of concern, including the modes (i.e. mechanisms) of contamination, for example, 'top down' spills, placement of fill, sub-surface release
- affected media, such as soil, sediment, groundwater, surface water, soil vapour and air quality (indoor air and ambient air), both on-site and offsite
- human and ecological receptors, both on-site and offsite
- complete exposure pathways, both on-site and offsite.

CSMs should logically explain the existing information, evidence, and data from the area under study, and be predictive. Where CSMs have poor predictive capabilities, the supporting information and evidence should be reviewed, and the CSM appropriately revised and updated, stating any data gaps.

Potential sources of contamination must be identified when investigating the site history, in accordance with EPA 2020b. After identifying the CoPC, their physico-chemical properties should be considered, such as solubility in water, volatility, miscibility and interactions with environmental media. Such consideration is especially important when dealing with uncommon or emerging contaminants, where behaviours, risks and remediation approaches may be less well known.

When conducting a preliminary site investigation (PSI), the available environmental information and site history information should be synthesised into a CSM. This preliminary CSM and any data gaps should feed logically into a SAQP. At every subsequent stage of site assessment, the CSM should be refined with the information and data from each investigation stage. Each refined CSM should be used to inform subsequent decisions on the condition of the site or area under study.

A CSM should identify uncertainties and data gaps relating to the contamination and potential exposure pathways. Any theories or assumptions underlying CSMs should be clearly identified to ensure adequate transparency. CSMs should address:

- how representative the available data are likely to be
- the potential sources of variability and uncertainty
- how important the identified gaps are to the objectives and reliability of the site assessment.

CSMs can take various forms, including text, tables, graphics and flow diagrams. They can also take the form of site-specific plans and figures, including cross-sections. The appropriate form of a CSM depends on a range of factors, including site complexity and the intended audience.

In developing the CSM, the consultant needs to distinguish between variability and uncertainty. Variability arises from diversity in the environment such as lateral variations in soil properties or lithology or changes in contaminant levels over time and space. Uncertainty represents lack of

knowledge about factors such as contaminant levels, which may be reduced after more investigation (NEPC 2013, B2).

While statistical analysis can provide a quantitative basis for decision making, the assessment of site contamination relies on multiple lines/weight of evidence including site histories, field samples, and geological and hydrogeological data and information. This approach allows scientifically defensible decision making supported by robust CSMs.

The site history and CSM should be developed in accordance with the requirements outlined in EPA 2020b.

2.2. Modes of contamination

When assessing site contamination, the mode of contamination affecting the site must be identified. The contamination's distribution is affected by the duration of the spill or leak; volume of contaminant lost; contaminant type, nature and age; the sub-surface material and whether preferential pathways are present.

Examples of modes of contamination include:

- filling or emplacement of materials from on-site or offsite areas with unknown contamination issues – examples include historical industrial waste from combustion furnaces or waste products, fill sourced from agricultural lands potentially contaminated with pesticides or other chemicals, building and demolition wastes or abandoned production materials
- heterogeneous filling from different unknown sources, which may result in the site varying in its spatial distribution of matrices and contaminant levels in ways that are not predictable; this is often referred to as "fill of unknown origin"
- top down contamination, where a leak or spill on the surface of the site filters down through the sub-surface – sources include above-ground tanks, drums, direct application of liquid wastes and spent liquors, transfer systems or vehicles
- sub-surface leaks, where contaminants leak from sub-surface infrastructure such as underground petroleum storage systems (UPSSs), trade-waste systems, septic tanks, sumps, pits, transfer lines or pipelines
- in-situ contamination, which is similar to sub-surface leaks but relates to contamination already in the sub-surface – examples include a leachate plume emanating from a landfill or contaminated soil, or phase-separated hydrocarbon (PSH) in the vadose zone above the water table, both of which can contaminate groundwater.

Some practitioners identify different modes of contamination as 'point sources' or 'distributed'. An example of a point source is a leak from an underground storage tank (UST), while an example of a distributed mode is fill of unknown origin. In addition, some naturally occurring materials can contain elevated levels of some contaminants, such as metals.

The modes of contamination should be considered in the investigation objectives and discussed in the CSM.

2.3. Data quality objectives

The DQOs process is used to develop performance and acceptance criteria (or data quality objectives) that clarify study objectives, define the appropriate type and quality of data, and specify tolerable levels of potential decision errors. These criteria are used as the basis for establishing the quality and quantity of data needed to support decisions. EPA policy is that DQOs must be adopted for all assessment and remediation programs, and that the process must be conducted before any investigative works begin (EPA 2017; NEPC 2013, B2).

Developed as part of the environmental data life-cycle process by the United States Environmental Protection Agency (USEPA), the seven-step DQOs process is a systematic planning method that

includes options for the type of problem to be addressed, based on the intended use of the data to be collected. The two primary types of intended use are classified as **decision making** and **estimation**.

The DQOs process, including the use of SAQPs, is further described in Appendix A: DQOs and the environmental data life-cycle process. Refer to USEPA 2000a, G-4HW and USEPA 2006b, G-4 for details of the process for collecting environmental data. Appendix B: Data-quality objectives: worked example gives a worked example of a hypothetical investigation-level decision problem.

Part 2 of these guidelines, *Sampling design part 2 – interpretation*, provides guidance on interpreting results.

3. Environmental sampling considerations

A population can be defined as any large collection of objects, things or individuals with some characteristics in common, that is being studied and for which information is sought. Generally, not all of a population can be measured, so a collection of measurements or observations is made as a sample of the population. The characteristics determined from the sample are then used to provide information about the population as a whole. In the assessment of site contamination, populations commonly include such things as the soil at a site or in a decision area, the fill in a stockpile, the gas in the soil, or the groundwater beneath the site.

The sampling of environmental media presents unique challenges for measurement due to matrix interferences, large-scale spatial variation, small-scale variations from matrix heterogeneity, temporal and seasonal variation in the target population, and the generally small number of measurements made relative to the media being assessed. Because of these factors, both multiple lines of evidence and weight of evidence approaches must be used in the assessment of site contamination, to synthesise the physical and numerical information that characterises a site and its surroundings. The CSM and the associated data-gap analysis are the key tools for this synthesis. In following this process it is imperative that the reporting includes the full physical and numerical dataset, and that methodologies are documented and explained, including any assumptions and associated limitations.

3.1. Types of samples

For the purposes of contaminated land assessment, a sample is usually a physical object: it can be a jar of soil, a cannister of soil gas, a bottle of water or an individual specimen of biota that can be chemically analysed at a laboratory for the CoPC. For soils the sample is quantifiable as a volumetric unit with physical, chemical, biological and spatial properties relevant to its source.

A sample does not have to be contained. Samples can be qualitative such as visual and olfactory observations, descriptions and field logging data which can be field-screened and then subject to other non-laboratory assessments and tests. A sample can also include Photoionization detector (PID) readings and soil gas readings obtained using handheld devices in the field. All the above examples are known as **field** samples.

Any sample that is sent to a laboratory to be analysed is known as an **analytical** sample. An analytical sample is a field sample, but a field sample may not necessarily be an analytical sample[1]. An analytical sample may be part of a larger field sample, and is typically destroyed in the process of analysis. In statistics, 'sample' is also used to mean **n**, the number of samples or individual measurements.

Statistical analysis and inference with prescribed error rates is done mainly with analytical samples. Under the multiple lines of evidence/weight of evidence approach, field samples are critical to inform the CSM and assist in defining the sources and pathways. The number and type of analytical samples is determined by the CSM, DQOs and statistical determinations, and the iterative nature of the process requires assigning an appropriate, but variable, weighting to the available evidence from both types of samples.

For information on the analytical methods for sample preparation of soils see Rayment & Lyons (2011).

---

[1] When field samples are collected, some may not be analysed immediately. Those samples must be analysed within the laboratory holding times for extraction and analysis for the various contaminants, or new samples may have to be collected.

4. Objectives of sampling programs

Clear definition of sampling objectives is essential to developing a sampling strategy, as this influences the sample types, the sampling pattern adopted, and the number of samples taken. Information is generally being sought regarding the type, location, extent and severity of the contamination, and where necessary, its comparison with relevant threshold values to enable effective decision-making.

The specific objectives of any sampling program will need to be defined on a case by case basis, depending on the project level objectives, the CSM, the media to be assessed and the stage of the project. NEPC 2013 states:

> *The purpose of site assessment is to determine whether site contamination poses an actual or potential risk to human health and the environment, either on or off the site, of sufficient magnitude to warrant remediation appropriate to the current or proposed land use.*

NEPC 2013 also notes that adequate site characterisation is the foundation for appropriate assessment of health and environmental risks associated with site contamination.

Materials which are intended for resource recovery, waste processing or disposal must be sampled and assessed as per the relevant legislation, including any relevant statutory instruments such as Resource Recovery Orders and Exemptions.

4.1. The process of assessing site contamination

The assessment of site contamination generally includes sequential stages of assessment and management, shown in Table 1, with the types of environmental sampling conducted at each stage. See Appendix A for the application of DQOs and SAQPs within the site assessment life cycle. See Section 5.2 for more details of 'probabilistic' and 'judgmental' sampling.

Table 1   Site contamination assessment investigation stages and associated sampling

Investigation stage	Type of sampling
Preliminary site investigation (PSI)	Sampling is not always required at this stage, and if performed is generally limited to judgmental sampling of soil, fill, and/or surface water. The PSI may be limited to a site inspection with field observations to verify desktop findings.
Detailed site investigation (DSI)	Both judgmental and probabilistic sampling are performed, commonly of soil, fill and groundwater, but sometimes also of soil gas, indoor air, ambient air, surface water and sediments.
Implementation of the remedial action plan (RAP)	Includes sampling for compliance monitoring, which is generally judgmental, and waste classification, which is probabilistic. Also includes investigations of unexpected finds uncovered during the physical works, which can include probabilistic and judgmental sampling.
Validation investigation	Conducted using probabilistic sampling for broad areas and judgmental sampling for validating hotspots, beneath former structures or within excavations, tank pits, trenches, etc. Can also include validation of continuous or batch remedial processes.

Ongoing monitoring (if required)	Targeted to specific locations such as sentinel groundwater wells or air monitoring in basements, as the extent and magnitude of contamination has been identified in a previous assessment stage. An ongoing monitoring program is developed with consideration to the CSM, and is site specific.

Specialist studies may also be required as part of the site contamination assessment process, for instance to provide data for human health or ecological risk assessments, assessment of the broader environment adjacent to and/or down-gradient from the site, and as part of the remedial design. While the assessment is usually represented as sequential steps, the steps often consist of multiple, overlapping investigations. For example, soil sampling can lead to further delineation of the extent of the contamination and potentially also groundwater sampling or soil gas sampling, which can lead to further soil sampling to close subsequently identified data gaps. See EPA 2020b for further guidance.

4.2. Characterisation and validation

Guidance has traditionally made a clear distinction between **characterisation** and **validation**. While this may be appropriate in some circumstances, there is no practical distinction between a final characterisation sample and a final, post-remediation, validation sample, when they are both taken as the final sample which concludes that a sample location is below any specific criterion or action level at the required confidence level. Accordingly, the required quality of both representativeness and usability for final characterisation samples and final validation samples should be identical.

Where assessment and remediation projects occur over extended periods of time, areas that were characterised as suitable for the proposed uses must be maintained throughout as being suitable. If subsequent uses occur which could cause contamination, such as stockpiling of potentially contaminated material, uncontrolled dumping of wastes, or ongoing industrial use, then further characterisation or validation will be required. Similarly, if there are significant information gaps in the site history between characterisation and proposed changes in site use, further characterisation or validation will be required.

4.3. Sampling objectives

Project objectives are broad: for example, to determine if a site is suitable for a specified land use. Sampling objectives, however, need to be very specific and concise, and should be defined as part of the DQOs process. They should be clearly documented and set out the media to be sampled, the CoPC and the principal study question, including the possible outcomes resulting from the study question.

As sampling objectives are situation-specific, it is not possible to be prescriptive about objectives and sampling designs. The typical objectives of a sampling design for a site contamination assessment may include:

- characterise the nature and extent of contamination at a site
- characterise soil, fill, stockpiles or waste materials for waste classification
- assess whether contamination levels exceed a criterion or action level
- determine the background condition of a specified media
- determine if contaminant concentrations significantly exceed background levels
- determine whether certain characteristics of two populations differ by some amount
- estimate certain population parameters such as mean, variation or the 95th (or greater) percentile

- identify the location of hotspots of a specified size, or provide evidence that they do not exist within specified confidence limits
- delineate groundwater or surface water plumes
- identify if a preferential pathway exists
- determine changes to contaminant levels over time
- monitor the effectiveness of a remediation technique
- determine if offsite impacts have occurred to any media
- determine if identified contaminants pose a human-health or ecological risk.

Analysis and interpretation of sampling data should be conducted in the context of the defined sampling objectives. More detail on the DQOs process is provided in Appendix A: DQOs and the environmental data life-cycle process, and a worked example is discussed in Appendix B: Data-quality objectives: worked example.

5. Sampling design

The two main categories of sampling design, probabilistic and judgmental, are discussed in Section 5.1. Broad sampling strategies which may be applied to all media are discussed in Section 5.2, with media-specific information in Sections 5.3 to 5.9.

5.1. Probabilistic and judgmental sampling design

There are two main categories of sampling design; **probabilistic** sampling and **judgmental** sampling.

A **probabilistic** sampling design uses random selection (that is, the different units in the population under study have an equal probability of being selected). This type of design, properly applied, results in unbiased and independent data. The advantages of probabilistic sampling designs are that they:

- enable statistical inferences to be made
- provide the ability to calculate uncertainty associated with estimates
- provide reproducible results within uncertainty limits
- produce decision error criteria that are incorporated into the interpretation and presented in results, usually as confidence statements.

However, for an optimal design using probabilistic sampling, an accurate CSM is required, including a clear definition of the population to be sampled.

Judgmental sampling, also called targeted sampling, means deciding where and/or when to collect the samples. It relies on good site histories and/or site features being clear and distinct. Judgmental sampling can be an efficient method for determining the areas of worst-case impacts, and is useful where the site history is inadequate or the features of concern are obscured or not discernible. However, judgmental sampling needs a high level of experience and expertise to choose the sampling locations and interpret the resulting data. If it is undertaken using an underdeveloped CSM, it provides poor quality data for site characterisation and should not be solely relied on.

Data collected using judgmental sampling are generally not suited for use in statistical determinations, as statistical determinations relating the sample data to the population parameter, such as estimating confidence intervals or conducting hypothesis tests, are only valid if the sample data are unbiased and independent.

If judgmental samples **are** used for statistical determinations, and they are targeted to areas of contamination such as fill material, stained and odorous soils or impacted groundwater, the resulting data will probably be biased upwards, meaning the site will be determined to be more contaminated than it is. If biased data is used for a statistical determination, it must be clearly documented, and ramifications and limitations must be identified and discussed. It is recommended that the results from judgmental sampling and probabilistic sampling are treated as two different populations.

5.2. Sampling strategies

Sampling strategies generally employed in the assessment of site contamination include:

- judgmental sampling (targeted)
- systematic sampling (probabilistic)
- random sampling (probabilistic)
- stratified sampling (targeted and/ or probabilistic).

Determining site contamination involves two main tasks: first, delineating the spatial properties of the environmental medium, stratum or decision area of concern, and second, characterising the physical properties and chemical concentrations of the CoPC for that medium, stratum or decision area. This characterisation may also have a time-dependent component, particularly for waters and air. This may be represented by a trend, such as reduction of source concentrations through natural degradation, or it may be seasonal (cyclic) or weather dependent. A combination of strategies is often used, for example, targeted sampling for known features and/or specific media, and systematic sampling to provide adequate site coverage and data for statistical inference.

To determine the number of samples required for site characterisation or site validation as a function of variance in the dataset and confidence levels, see Section 7.

5.2.1. Judgmental sampling

Judgmental sampling is also called targeted sampling.

Judgmental sampling locations are selected based on the investigator's knowledge of the probable distribution of contaminants at the site, with known or suspected areas of contamination being specifically targeted based on the CSM. It is an efficient sampling method that makes use of site history and field observations, but is statistically biased. The quality of the resulting data depends in part on the experience and judgment of the consultant, and the available site history information and observable site features.

Judgmental sampling can also result in an uneven distribution of sampling locations, in which case additional sampling locations are required to provide site coverage. Judgmental sampling should not be used as the only method for site characterisation unless there is detailed and accurate documentation of the history and site information that can be provided to support the decision. This should be reflected in the DQOs.

Judgmental sampling is recommended to validate the remediation of solid media and the removal of infrastructure such as underground petroleum storage systems (UPSSs). The number of judgmental samples taken is determined in part by the number and size of potential identified sources, and the number and area of observable features, such as staining, odours, wastes and extent of fill material. Sample locations may also be targeted along potential migration routes of surface drainage or permeable materials.

Judgmental sample results should be reported in separate results tables from other results. If they are reported in the same table they must be clearly marked as belonging to a judgmental rather than probabilistic data set, to ensure appropriate data segregation.

As contamination concentrations can vary greatly over short distances, single judgmental samples may not provide a complete understanding of the potential contaminant range. Where areas of contamination are identified, further 'step-out' and depth sampling is often required, to determine if the likely maximum contaminant levels have been identified. Further sampling should identify if the concentrations are increasing or decreasing away from the identified contamination, noting that sharply defined boundaries rarely exist.

The use of judgmental sampling designs is often required for environmental media other than soil, as it is not always possible to sample these media at random locations and times. For groundwater, soil vapour and ground gas studies, consultants should seek to target all the relevant potential sources, receptors and pathways. Random sampling locations, while being more defensible statistically, essentially serve no purpose unless a large and generally cost-prohibitive number of samples is taken. See DEC 2007 and EPA 2020b for further guidance.

For surface waters, sampling often targets specific locations such as upstream and downstream of a site on a watercourse. Unless the watercourse or body is particularly large, randomising sample locations is impractical. Generally, there is a high degree of natural mixing and homogeneity in surface waters, but stratification can occur. Sampling design should ensure appropriate controls, to

minimise unrepresentative samples. Refer to ANZG 2018 for more information on sampling surface waters and sediments.

Figure 1 shows an example of judgmental sampling, with sampling locations (blue dots) concentrated around a potentially contaminating underground storage tank (grey box).

Figure 1   Judgmental sampling example 1

● Sampling point location
▢ Underground storage tank
----- Gridded lines providing a guide of relative placement of sampling point locations
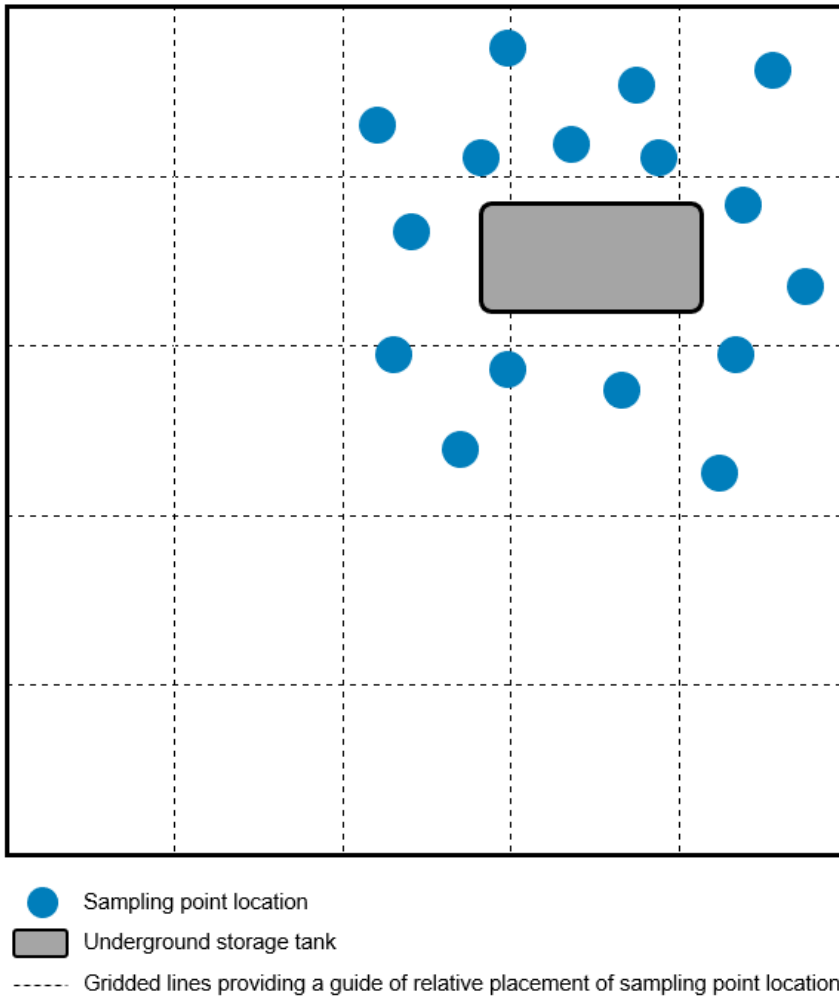
Source: NSW EPA

Figure 2 is an example of judgmental sampling showing sampling locations (blue dots) around a contaminated underground storage tank (grey box). When contamination is detected above assessment criteria (locations shown as blue dots with red crosses), subsequent sampling locations have been used to delineate the contamination (green dots).

Figure 2   Judgmental sampling example 2
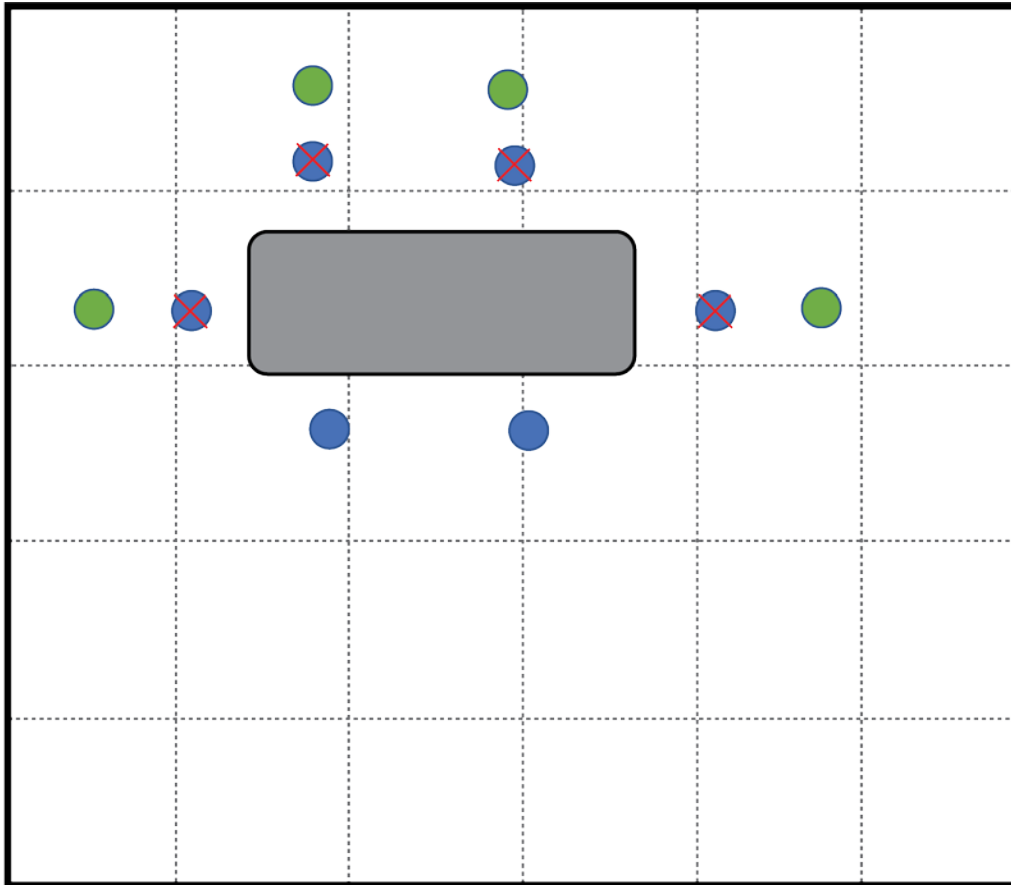
- Initial sampling locations around known location of an underground storage tank.
- Initial sampling locations found to be higher than criteria.
- Subsequent "step out" sampling locations around the underground storage tank.
- Underground storage tank
- Gridded lines providing a guide of relative placement of sampling point locations.

Source: JBS&G

5.2.2. Systematic sampling

Systematic sampling is a probabilistic strategy that involves selecting points at regular intervals over an area, for example, grid intersections, or time. Systematic sampling does not generate clusters of sampling locations but ensures an even coverage of the site or decision area, which makes this approach ideal for characterising sites or decision areas. Systematic sampling is statistically unbiased as long as the coordinates of the first sampling location are determined randomly.

Examples of systematic grids include square, off-set square, rectangular, triangular, herringbone (recommended in TCRBE 1994) and radial grids (NEPC 2013, B2). Square grids are generally used as they are simple to establish. They have also given adequate results (BSI 10175:2013) in studies that have evaluated the relative efficiencies of various systematic sampling patterns for hotspots of different shapes. For rectangular or triangular grids, consult Gilbert 1987.

In the assessment of site contamination, systematic sampling is usually done over a grid, although transects may be appropriate when lineal features are being assessed, such as the validation of former pipeline trenches. Gilbert 1987 notes that uniform coverage in many cases yields more accurate critical parameters of a contaminant distribution, such as the mean. NEPC 2013, B2 states that "systematic and grid sampling is used to search for hotspots and to infer means, percentiles or other parameters".

Generally, a convenient site feature or boundary is selected to establish a grid. The remaining sampling locations are then defined so all locations are at regular intervals over an area (NEPC 2013, B2). For regular square grids, the required grid size is set out, and samples are taken at the same location from each cell, ideally the centre of the cell. Alternatively, the sample-location coordinates within each cell can be selected using a randomised offset, between zero and the grid cell size in each dimension (see Figure 3). This design can be more practical for operational sites or where significant infrastructure exists, where selected locations are blocked.

The mesh size (the dimensions of the grid cells in both the x and y direction) is related to the size of a hotspot and the required probability of detecting a hotspot of a specified size. Where elongated hotspots are expected, possibly due to land slope, differential x and y mesh sizes will help to detect them. Appendix C: Determining sampling grids for hotspot detection gives a method for calculating grid size. Section 5.2.5. provides the minimum recommended number of samples for systematic sampling.

Figure 3 shows an example of a systematic sampling pattern, with sampling locations (blue dots) placed at regular intervals (indicated by equally sized squares marked by dotted lines) in a square grid across the investigation area.

Figure 3   Systematic sampling pattern example
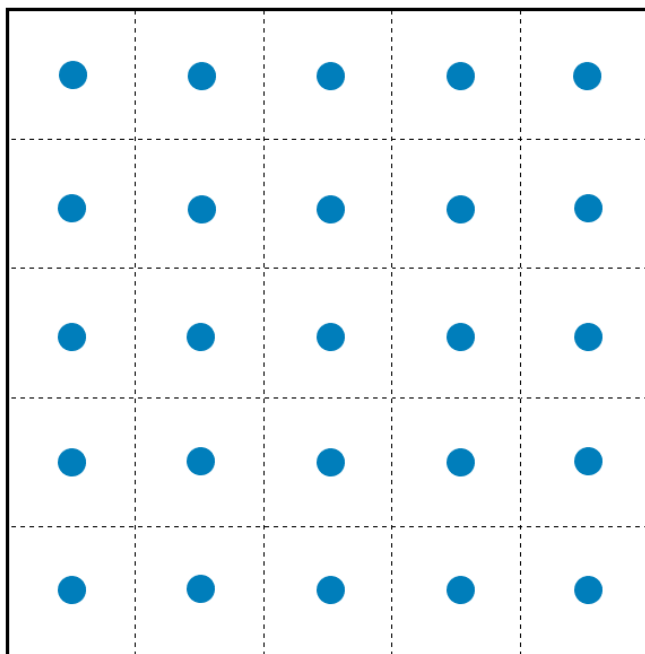
● Sampling point location

----- Gridded lines providing a guide of relative placement of sampling point locations

Source: NSW EPA

5.2.3. Random sampling

With random sampling, sampling points are selected randomly but not arbitrarily[2]. A legitimate uniform pseudo-random number generator, for example a computer program, should be used to determine sampling location coordinates. The randomisation process ensures unbiased data, as any location within the sampling area has an equal chance of being selected as a sampling point.

NEPC 2013, B2 states that simple random sampling 'is most useful when the area of interest is relatively homogenous, and no major pattern or hotspots are expected'. Examples may include specific decision areas for which no information is available, as part of a PSI. Where used in the assessment of site contamination, the limitations of random sampling should be considered and appropriately documented.

While random sampling is statistically unbiased, sampling points can cluster together by chance. This makes them deficient for detecting hotspots and for giving an overall picture of the spatial distribution of the contamination. In practice, random sampling has limited use in the assessment of site contamination, unless combined with systematic grid sampling (discussed in Section 5.2.2).

Figure 4 below shows an example of a systematic sampling pattern with a randomised offset, with sampling locations (blue dots) placed at random locations in each grid across the investigation area.

Figure 4   Example of a systematic sampling pattern with a randomised offset

●   Sampling point location

----   Gridded lines providing a guide of relative placement of sampling point locations
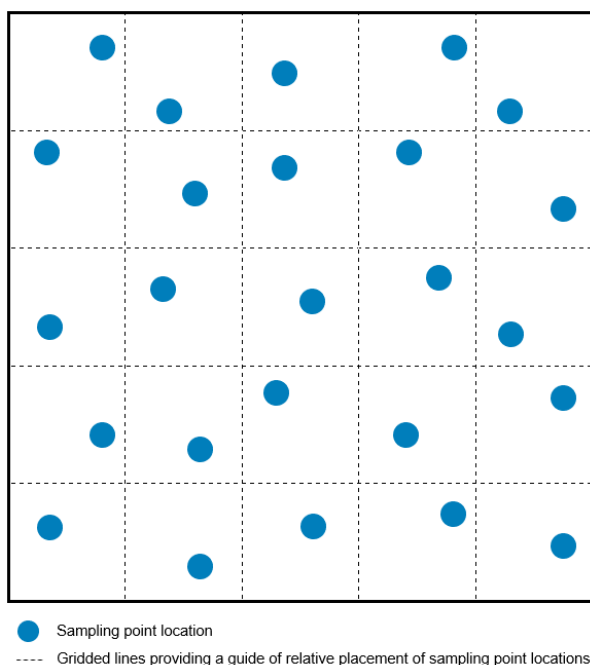
Source: NSW EPA

5.2.4. Stratified sampling

Stratified sampling may be appropriate for investigations of large sites with different uses, features and complex contaminant distributions. Under this approach, the site is divided into various non-overlapping sub-areas, according to geological and geographical features, the nature of the contamination, former usage of the site or other relevant factors. Each sub-area can then be treated as an individual decision area with different sampling patterns and sampling densities

---

[2] Arbitrary samples are also considered judgmental samples, as it is not possible to rule out unconscious bias.

applied. For example, one area might require a targeted sampling strategy while a neighbouring one needs a systematic strategy.

A stratified sampling strategy requires reliable prior knowledge of the site. In some cases, stratified sampling patterns may require more complex statistical analysis, as discussed in Gilbert 1987 and USEPA 2006a, G–9S.

NEPC 2013, B2 describes the following advantages of implementing a stratified sampling pattern:

- potential for achieving greater precision in estimates of the mean and variance where the measurement of interest is strongly correlated with the variable used to define the strata
- calculation of reliable estimates for subgroups of special interest.

Figure 5 is an example of a stratified sampling pattern, with three separate investigation areas, sampling strategies and sampling locations due to different characteristics of the site. Area 1 uses a high-density judgmental sampling strategy with many sampling locations (blue dots) clustered around an underground storage tank (grey box). Area 2 uses a medium-density systematic sampling strategy to assess fill material from an unknown source, with sampling locations (orange dots) at regular intervals in a grid. Area 3 uses a low-density systematic sampling strategy with sampling locations (green dots) further apart than those in area 2, to assess natural soil with no known contaminants. Data will be analysed as three different data sets.

Figure 5   Stratified sampling pattern example
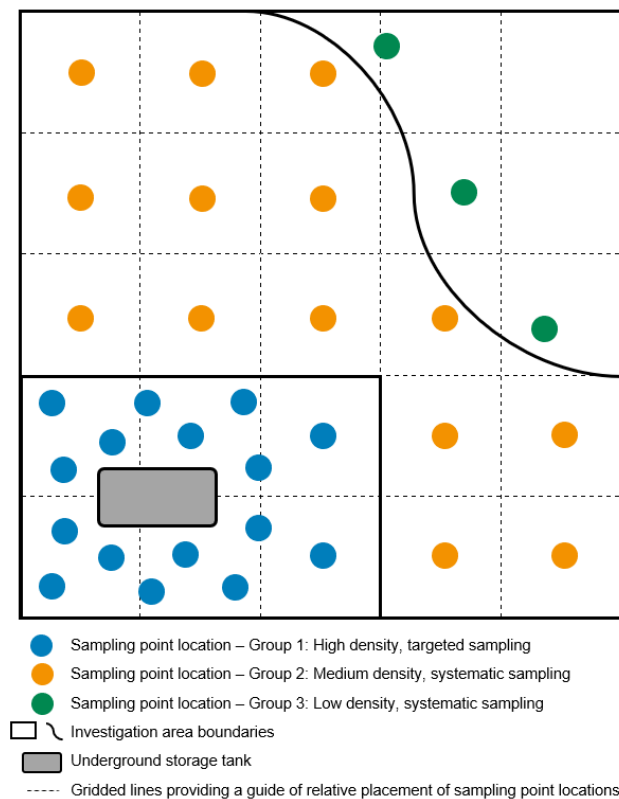
Source: NSW EPA

See Appendix H for a worked example using a stratified sampling regime.

5.2.5. Minimum recommended number of samples for systematic sampling

The preliminary CSM should guide the development of an appropriate sampling density, and a sampling grid size should be calculated based on the size of hotspot that the site history suggests is likely to be present at the site (see Section 5.2.1 and Appendix C: Determining sampling grids for hotspot detection).

Site histories can be incomplete and the locations of buried contaminant sources, such as drums, underground tanks, livestock dips or animal carcasses, might not appear on historical records. In particular, fill of unknown origin can be a source of contamination that can only be identified by intrusive investigations supported by the collection of analytical samples.

When determining the appropriate number of sampling locations to account for the lack of, or uncertainties in, site history information, or when fill of unknown origin is suspected or known to be on the site, using a regular square-grid systematic pattern is recommended. This approach should produce an initial data set with enough results to enable some statistical parameters to be determined, and satisfies the following acceptance criteria:

- the arithmetic average concentration of the contaminant(s) must be less than an acceptable limit, at a 95% or higher confidence level

- a site must be free of hotspots larger than a critical size, at a 95% or higher confidence level.

The formula for calculating the 95% upper confidence level of the arithmetic mean can be found in appendices J, K, and L in Part 2 of these guidelines.

Table 2 provides recommended minimum sampling densities for site characterisation based on detecting circular hotspots using a square-grid, systematic sampling pattern. The recommended number of sample locations is a minimum only, and should not be considered appropriate in all circumstances.

For sites more than five hectares in size, it is recommended that the site be subdivided into smaller areas for effective sampling.

If there are any areas of concern, such as dead vegetation, structures, or evidence of disturbed ground, the area should be separated and subject to stratified sampling as discussed in Section 5.2.4. If the site history indicates that potentially contaminating activities have been carried out at the site, then targeted sampling should be performed in the vicinity of potentially contaminating activities, as well as systematic sampling performed on a grid, using a higher sampling density than the one in Table 2.

For systematic sampling, any use of lower sampling densities than those listed in Table 2 must be accompanied by sufficient justification in both the SAQP and Detailed Site Investigation (DSI), including confirmatory samples and observations made during a site walkover.

Further sampling may be required for delineation or validation purposes.

Table 2   Minimum number of sampling points for a square grid, based on site area

Size of site (ha)	Minimum number of sampling locations	Grid size (m)	Diameter of the hotspot that can be detected with 95% confidence (m)
0.05	8	8	9.3
0.1	8	11	13.2
0.2	8	16	18.7
0.3	9	18	21.5
0.4	11	19	22.5

0.5	13	20	23.1
0.6	15	20	23.6
0.7	17	20	23.9
0.8	19	21	24.2
0.9	20	21	25.0
1.0	21	22	25.7
1.5	25	24	28.9
2.0	30	26	30.5
2.5	35	27	31.5
3.0	40	27	32.3
3.5	45	28	32.9
4.0	50	28	33.4
4.5	52	29	34.7
5.0	55	30	35.6

Greater sampling densities and sample volumes should generally be used for asbestos than those considered appropriate for other contaminants. This is because asbestos can occur widely and unpredictably and, as a discrete contaminant, it can be hard to detect using conventional sampling regimes (WA DOH 2009).

For more information, see:

- Appendix C: Determining sampling grids for hotspot detection, for the number of samples needed to detect hotspots of given sizes
- Section 7 for how to calculate the number of samples required to be representative of a population
- Appendix E: Determining the number of samples by the CRV method, for determining the number of samples by the CRV method
- Appendix F: Determining the number of samples by the MPE method, for determining the number of samples by the MPE method.

5.3. Soil and fill material

'Soil' describes the naturally occurring or residual soil that forms due to weathering or geomorphological processes. Soil is the primary medium of concern when assessing site contamination, and has traditionally been the focus of the assessment and management of contaminated land.

While generally homogenous in the absence of anthropogenic contaminants, Hamon et al. 2004 note that trace elements have naturally high variability. Depending on the associated parent material, metals can be highly variable: for example, ultra-mafic rocks can lead to naturally elevated chromium (Cr) and nickel (Ni) levels. It is recommended that the soil landscape mapping be interpreted for variability so background soil heavy metal concentrations can be assessed. Alluvial and estuarine geomorphic systems can display very high variability due to the mobility of waterways in the environment and shifts in drought and flood regimes over time. This resulting deposition can lead to the formation of unique geological features in very small areas of the soil.

Emplaced fill or 'made ground' refers to excavated earthen materials or wastes that have been deposited on a site by artificial means, often to build up or level the surface of a site. Depending on the site, its location and when the site was filled, the material may consist of fill from on-site cut-and-fill activities; overburden material received from offsite locations; industrial wastes such as furnace wastes, ash, slags and tailings; construction and demolition wastes; biosolids; and other industrial wastes and residues.

The potential for fill to be present at a site should be identified during the PSI and detailed in the CSM as a potential contaminant source.

While some monolithic deposits can be highly homogenous, fill is often highly heterogeneous. Each fill layer must be sampled discretely and separately from underlying natural materials. This is because both fill and natural soils should be sampled as part of site characterisation, with care taken to collect discrete samples from each specific target stratum. Identified fill material should be appropriately described and logged during site investigations, and data analysis should be conducted by material type, rather than different soils and fills being analysed as one material, which can result in interpretation errors.

To ensure the representativeness of samples, assessment of fill of unknown origin should preferably use test pits to provide a larger exposure of the fill layer, so the small-scale variability of fill is recognised and inspection for Asbestos Containing Material (ACM) is facilitated.

5.3.1. Depth of sampling

The sampling depth and interval is dependent on the CSM, DQOs and mode of contamination. NEPC 2013, B2 states that 'at the surface, samples at 0–100 mm or 0–150 mm should be taken unless there is evidence of a thin surficial layer of contamination'. Examples of such situations include rifle ranges and broadacre agricultural sites, where some analytes tend to accumulate in a thin surficial layer, and areas that have received surface applications, such as termiticide sprays beneath slabs. Samples should be collected from both the emplaced fill and natural soils, at intervals of generally no more than 500 mm, and at locations where distinct differences in permeability or other observable features occur, as per NEPC 2013, B2.

The following should be considered when deciding on the sampling depth interval:

- the likely fate and transport of the CoPC
- whether permeable layers are present in fill and natural soils
- the mode of contamination
- visual/olfactory indicators of contamination (in these cases, the use of field screening tools such as PIDs can help identify the depth of sampling).

A sample should be collected beneath the point where fill meets the underlying natural soil.

If validating potential impacts from subsurface infrastructure, samples should be at a depth that is likely to intercept any associated contamination (e.g. if underground tanks have been removed, samples should be collected from the lower half of the excavation wall).

Constituent samples should be from the smallest depth interval consistent with providing adequate representation of the interval (Standards Australia 2005).

5.4. Stockpiles

5.4.1. Stockpile sampling principles

It is preferable to characterise soil and fill *in situ*, but at times site or project-specific constraints require material to be stockpiled before it is sampled. The excavation and stockpiling of material can result in the mixing and dilution of contaminated materials with uncontaminated materials. The excavation and placement of the material should therefore be supervised to ensure different types

of soil and fill materials are kept segregated. For example, topsoil or fill should be stockpiled separately from underlying natural material, and discernibly different types of fill should be stockpiled separately.

Where stockpiles are pre-existing, site investigations should ensure they are fully examined and that samples are collected from the entire stockpile, not just the surface. This can involve the use of excavators, drill rigs or hand augers to help access the interior of the stockpile.

The history of a stockpile may be very different to the site history of the land on which the stockpile is located. Therefore, all stockpiles on a site should be inspected and sampled. Photographs and recorded field observations can assist with record keeping, especially if stockpiles have been left unattended.

A sampling strategy for a stockpile should consider its observable composition, history, size and the proposed end use of the material. The sampling strategy should also consider that a stockpile is three dimensional, and requires systematic sampling, using three-dimensional grids for characterisation (EPA Victoria 2009).

Field sample descriptions should include comments on the presence or absence of visual and olfactory contamination. The results of PID characterisation and analytical sample results should be tabulated against relevant criteria.

The coefficient of variance (CV) can be calculated to assist in the interpretation of results from stockpiles, particularly where there is a large range in values and a need to determine if the data is homogenous. See the Glossary for more information.

5.4.2. Purpose of stockpile sampling

There are several scenarios where stockpiles on a site may require sampling, including:

1. Sampling for on-site reuse
2. Sampling for offsite reuse
3. Sampling for offsite disposal to landfill or transport to a recycling facility.

The purpose of sampling stockpiles should be determined first, as specific sampling requirements apply to each of the above scenarios. These are discussed in Sections 5.4.3 to 5.4.5.

In some instances, the fate of a stockpile may not be determined until after the results of sampling and analysis are obtained, and additional sampling may then be required.

See Appendix I for a flow chart outlining the process of stockpile assessment for stockpiles not impacted by asbestos.

5.4.3. Stockpile sampling for on-site reuse

If material is intended to be kept and used on-site, it must be demonstrated that the material is suitable for use at that site. For imported materials, additional sampling and analysis beyond what was undertaken to comply with resource recovery orders and exemptions may be required to demonstrate suitability of use. Stockpile sampling for on-site reuse should consider Table 3 or Table 4.

For sampling requirements for stockpiles suspected of being impacted by asbestos, see Section 5.4.7.

For material being retained for use at the site, the number of samples required for sampling a stockpile can be derived using the methods described in Section 7, that is, the combined risk value (CRV) method (see Appendix E: Determining the number of samples by the CRV method) and the maximum probable error (MPE) method (see Appendix F: Determining the number of samples by the MPE method).

5.4.4. Stockpile sampling for offsite reuse

Sampling of stockpiles for offsite reuse must comply with an appropriate Resource Recovery Order and Resource Recovery Exemption, or must comply with the definition of Virgin Excavated Natural Material (VENM) set out in Schedule 1 of the *Protection of the Environment Operations Act 1997* (POEO Act).

Resource Recovery Orders contain specific requirements that must be complied with for the purposes of supply for offsite reuse. This includes requirements about the number of samples to be collected and chemicals and attributes to be tested for, amongst other requirements. For more information about orders and exemptions, refer to the EPA website.

Table 3 or Table 4 can assist in classifying VENM.

5.4.5. Stockpile sampling for offsite disposal to landfill or transport to a recycling facility

Stockpiles of material that require off-site disposal to landfill must be classified according to EPA 2014c and other EPA approved guidelines. Stockpiles of material that require waste classification for disposal or transport to a recycling facility must be sampled in accordance with the minimum number of samples outlined in Table 3 and Table 4.

For sampling requirements for stockpiles suspected of being impacted by asbestos, see section 5.4.7.

5.4.6. Minimum number of samples recommended for stockpile sampling

Table 3 provides the minimum number of samples recommended for characterisation of stockpiles containing up to 200 m^3 of homogenous soils. Greater sampling densities are required for stockpiles that contain heterogenous material or have large ranges in contaminant concentrations.

Where there is a large range in contaminant concentration, either the maximum concentration should be assumed for management purposes, including for disposal, or additional samples should be collected and analysed and the situation re-evaluated (NEPC 2013, B2). Different sampling rates may be appropriate for soil quantities of more than 200 m^3. Statistical methods to apply in this situation are discussed in Section 7.

Table 3   **Minimum number of samples recommended for initial assessment of stockpiles of up to 200 m^3**

Stockpile volume (m^3)	No. of samples
<75	3
75 – <100	4
100 – <125	5
125 – <150	6
150 – <175	7
175 – <200	8

Table modified from NEPC 2013, B2

Two approaches can be taken for assessing stockpiles of more than 200 m^3:

1. Samples can be collected at a rate of 1 sample per 25 m^3
    or
2. Samples can be collected at a reduced frequency, subject to a 95% UCL calculated for all applicable analytes, and this value is then compared with the relevant assessment criteria. This

method is not appropriate for sampling stockpiles impacted, or potentially impacted, by asbestos.

These approaches are summarised in Table 4.

Table 4   **Minimum number of samples recommended for initial assessment of stockpiles over 200 m³**

Stockpile volume (m³)	No. of samples (1:25 m³)	Minimum number of samples for 95% UCL (not for asbestos)
200 - 300	12	10
400	16	10
500	20	10
600	24	10
700	28	10
800	32	10
900	36	10
1000	40	10
1500	60	10
2000	80	10
2500	100	10
3000	120	12 (1:250)
4000	160	16 (1:250)
4500	180	18 (1:250)
5000	200	20 (1:250)
>5000	1:25	1:250

Table modified from Vic EPA 2009

5.4.7. Sampling for stockpiles suspected of being impacted by asbestos

If a stockpile is suspected of being impacted by asbestos, sampling for asbestos must be undertaken to confirm whether or not asbestos is present. Reference should be made to the CSM, site history, local knowledge, and the presence of uncontrolled fill or building materials.

Sampling for stockpiles suspected of being impacted by asbestos must be undertaken in accordance with Table 5.

Table 5   **Minimum number of samples for stockpiles suspected of being impacted by asbestos**

Purpose of sampling	No. of samples required	Requirement
On-site reuse	Table 4 should be considered	Prior to reuse, stockpiles must be validated as suitable for the approved land use.
		Comply with guidelines approved under the *Contaminated Land Management Act 1997*, including the Site Auditor Guidelines, and relevant provisions of the POEO Act, including s 144AAB, where applicable.
Offsite reuse	Refer to EPA website for relevant Resource Recovery Order and	Comply with the relevant Resource Recovery Order and exemption.

	exemption that may prescribe sampling and other requirements for asbestos. Table 3 or Table 4 may be used to assist in classifying VENM.	Comply with VENM requirements if applicable.
Disposal to landfill	3 samples for stockpiles less than 75 m³, plus 1 sample for every additional 75 m³	Comply with EPA 2014c, and take to a disposal facility lawfully able to receive the waste for disposal.
Transport to recycling facility	Table 4 must be complied with	Test each sampling location for: • non-friable asbestos using the NEPM gravimetric procedure where the sample volume must be a minimum of 10 L per sample, and • asbestos fines/ fibrous asbestos ('AF/FA') where the sample collected must be a minimum of 500 mL. Comply with EPA 2014c, and if no asbestos is present, take to a recycling facility lawfully able to receive the waste for recycling. If a stockpile contains asbestos it must not be recycled.

Source: NSW EPA

Refer to the EPA website for further guidance about the lawful management of asbestos-impacted soils and stockpiles.

5.5. Validation

An SAQP should be developed for validation, with validation samples collected on a systematic grid. The optimal number of samples can be determined using the CRV or the MPE methods (see Section 7) and laid out using a systematic grid as described in Section 5.2.2.

For excavations, at least one validation sample should be collected from the bottom and from each of the pit walls. For large excavations, a sampling grid should be established based on field observations and the CSM for the site.

For instance, validation of an excavation at a former shooting range that is impacted by lead pellets could require validation on a small grid of 5 m, due to the scattered nature of the source of the CoPC. Validation of an excavation following a UST exhumation could require one or two samples collected every 10 lineal metres along the wall and a sample collected for every 25 square metres of the base. This will depend on the CSM for the site, and the rationale should be clearly documented in the RAP.

For the validation of continuous remedial processes, an SAQP should be developed based on the remedial methods and the CSM.

5.6. Use of composite samples

Composite sampling of soils involves mixing several discrete samples or sub-samples of soil to form one composite sample for analysis. Composite samples should only be used in former orchards and market gardens as described in DEC 2005a.

The maximum number of discrete samples that are allowed is four (NEPC 2013). To ensure they are representative of similar materials, samples must be collected from the same stratigraphic unit and from no further apart than 20 m. Subsamples for compositing should not be collected where there is spatial or temporal variability.

Composite samples cannot be used for validation purposes (AS 4482.1-2005).

In principle, the concentration of the composite sample represents the average of the sub-samples. However, composite sampling has three major drawbacks:

- It cannot be used to assess pH, or volatile or semi-volatile contaminants including TRH, BTEXN, OCPs, OPPs and low molecular weight PAHs. As a result, a good understanding of the site history and the CoPC are necessary for adopting a composite sampling approach (NEPC 2013, B2).
- Composite sampling is not suitable for clay or fine-grained soils, as subsamples are difficult to mix adequately.
- A simplistic analysis of composited samples can result in a sub-sample that contains a high concentration of contaminant, which can remain undetected due to the dilution effect of the compositing process, potentially resulting in unrepresentative data and associated decision errors.

Where composite sampling has been used, the relevant assessment level should be divided by the number of sub-samples in the composite and compared with the laboratory result. (NEPC 2013, B2).

Further information about composite samples can be found in the NEPC 2013, B2 and DEC 2005a.

5.7. Groundwater

Potential groundwater contamination must be considered when designing sampling programs at contaminated sites. These design requirements are impacted by the type and nature of the site's groundwater system, which can be complex and have multiple interacting aquifers.

The appropriate method for the assessment of groundwater is determined by undertaking a PSI. This should include a desktop hydrogeological assessment and a site-specific CSM, which must include groundwater. To inform the CSM, published geological reports and hydrogeological information for the surrounding area can be found on NSW government websites. In conjunction with field observations and soil analytical results, the geological information can help determine the number and location of groundwater wells, screen intervals and well depths. If published information does not accurately represent the conditions encountered in the field, well locations, screen intervals and well depths should be determined to suit the encountered field conditions.

Groundwater wells are generally installed in locations that will maximise the likelihood of intercepting and defining the extent of groundwater contamination and evaluating the potential for off-site migration, or potential pathways to an on-site receptor. This should include targeting contamination sources and known plumes, and then locating wells hydraulically up-gradient, down-gradient and cross-gradient (lateral) to the areas of concern. If a potential for offsite migration exists, groundwater monitoring wells should be installed as close to the down-gradient site boundary as practical. The location of down-gradient groundwater receptors should also be considered. Wells must be installed so the groundwater flow direction can be determined.

The screens in the wells need to target the appropriate aquifer/water-bearing zone or the zone of interest in that aquifer. The physical and chemical characteristics of the contaminant can affect their distribution in groundwater. Multiple wells or wells with multiple screens/nested wells may be required to characterise the vertical groundwater profile and contaminant distribution. Fluctuations in groundwater level due to tidal influence, seasonal conditions and groundwater extraction should also be considered.

There are various sampling methods for the collection of groundwater samples. Site-specific conditions and the CoPC must be considered when selecting an appropriate method.

High-resolution site characterisation techniques can also be deployed to characterise contamination, depending on the CoPC. Details can be found on the ITRC website, Implementing Advanced Site Characterization Tools (itrcweb.org) and in NEPC 2013, section 7.2, B2.

The information in this document should not be wholly relied on when designing a groundwater monitoring program. Specific guidance on groundwater sampling design and sample collection can be found in NEPC 2013, B2 and DEC 2007. Methods for the statistical analysis of groundwater data, including intra well and inter well comparisons, can be found in USEPA 2009.

5.8. Surface water

The sampling design for a surface water program should take into account the CSM, the purpose and objective of the program, the chemical characteristic of the contaminants and the pathways and receptors.

Consult the ANZG 2018 website before designing any surface water monitoring program.

5.9. Sediment

Sediment is unconsolidated mineral and organic particulate material that has settled to the bottom of aquatic environments. Sampling of sediments should be undertaken with reference to ANZG 2018 and Simpson & Batley 2016. Material which has been removed from the aquatic environment and forms part of the land should be assessed as soil.

5.10. Vapour

For vapour investigations, multiple lines of evidence should be used. The CSM must include:

- the design and condition of existing or proposed buildings, including the presence of elevators and ventilation systems
- preferential pathways – both constructed pathways (such as building sumps, drains, services and permeable backfill) and natural pathways (such as tree roots, differential soil permeabilities and fractured bedrock)
- environmental factors such as diurnal fluctuations, short-term and seasonal fluctuations in weather conditions, and variations in soil moisture and temperature
- confounding sources of contamination that may contribute to the volatile organic compounds (VOCs) measured at the sites, for both indoor and ambient air.

Vapour sampling should be undertaken with reference to NEPC 2013, EPA 2020a and CRC Care 2013.

5.10.1. Soil vapour

Soil vapour sampling is generally the preferred approach where vapour from a sub-surface source is likely. The number of sample locations recommended spatially for a vapour investigation depends on site-specific conditions. Access constraints such as building construction and occupation can significantly impact sample locations for soil vapour assessments, and different types of samples such as indoor air may be needed to obtain a weight of evidence approach.

Soil vapour sampling should target the highest concentrations, either known or expected, at the site, and the location of current or future receptors (that is, inhabited buildings or the location of a proposed building). Additional samples should be collected between the source of contamination and all potential receptors.

Refer to EPA 2020b and CRC Care 2013 for further information on vapour intrusion and soil vapours (trace ground gases).

Various sampling methods are available for collecting soil gas samples, including active and passive methods. Site-specific conditions, the CSM and the contaminants of concern must be considered when selecting an appropriate method.

Sample frequency

At least one round of sampling should be taken in weather conditions that are likely to result in the highest vapour concentrations, considering temperature, pressure, and soil moisture. There should be repeat sampling where site conditions may change – for example, there is a fluctuating source, varying meteorological conditions, varying building use or conditions, or remedial work is being undertaken. If undertaking a health risk assessment for vapour intrusion, additional sampling rounds should be used to address the potential for variability of the contaminant concentrations. CRC Care 2013 identifies that there is no need to repeat sampling if soil gas values are a factor of 5–10 times below the risk-based screening levels, unless there is a major change in conditions (such as an elevated water table) that would significantly change vapour concentrations.

5.10.2. Indoor and ambient air

Indoor air sampling is the most direct method of sampling VOC exposure where the CSM has identified that vapour intrusion is a potential pathway. Where concentrations of CoPC in indoor air attributed to a source of contamination exceed relevant criteria, the appropriate parties should be notified and the need for mitigation measures should be assessed.

The number of samples recommended for representative indoor air sampling depends on the size of the indoor area and the building's internal divisions, which may limit air movement. Air samples should be obtained from the crawl space and/or basement if present, and the living area at the height where occupants sit or sleep. Overall, sample locations should be targeted to inhabited buildings, with samples collected from where people will be breathing.

Sources of VOCs inside buildings should be considered before sampling and assessing if indoor air sampling is appropriate in the context of a weight of evidence approach. Examples of indoor VOC sources may include aerosol cans, petrol and other light fuels, dry cleaning chemicals, solvents used for cleaning, new and near-new building materials, floor coverings and furniture.

5.10.3. Ground gases

Ground gases may have natural or anthropogenic origins, and can be generated as a result of the biological, chemical and physical decomposition of landfilled, spilled or dumped wastes. They may be associated with old mine workings, operating or closed landfills, coal and peaty soils or buried putrescible wastes. The assessment of ground gases is a complicated area of investigation and is beyond the scope of these sampling design guidelines: see the specific information in EPA 2020a and its references.

5.11. Determining background concentrations

Knowledge of the background concentrations of an analyte at a site is important in understanding how much contamination may be present, particularly when assessing metals and metalloids. Metals and metalloids are naturally occurring elements: their natural 'background' concentrations in soils are highly variable, and depend on the rocks from which the soils originate and the processes occurring during soil formation (Gray & Murphy 1999).

Arguably, as a consequence of the industrial revolution, natural background concentrations no longer exist, at least in surficial soils, due to anthropogenic sources and the global transportation of contaminants. NEPC 2013, B5b states that 'the term ambient background concentration (ABC) … is used rather than background concentration'.

NEPC 2013, B1 assumes that ecosystems are adapted to the ABCs of metals in soils, and only the addition of contaminants above this background concentration has an adverse effect on the environment. It notes that:

The ABC of a contaminant is the soil concentration in a specified locality that is the sum of the naturally occurring background level and the contaminant levels that have been introduced from diffuse or non-point sources by general anthropogenic activity not attributed to industrial, commercial, or agricultural activities, for example, motor vehicle emissions.

This definition can be extended to other media.

Determining the ambient concentrations for any medium relies on identifying sites or decision areas that have not been affected by the same or similar contaminating activities as the subject site or decision area (or, if that is not possible, not affected to the same magnitude).

The following should be considered when determining the ABCs of various media.

5.11.1. Soils

- Ensure that the background areas consist of the same soil types as the site or decision area.
- Collect and compare samples with soils and sediments from the same soil horizon layer.

5.11.2. Groundwater

- Construct background groundwater monitoring wells in the same way as the subject site wells, targeting the same aquifer.
- Consider potential sources of contamination up-gradient of the well.
- Assess preceding rainfall and standing water levels.
- Collect and record physico-chemical parameters at both the decision site and an unaffected site.
- In highly fractured or karstic geological environments, seek specialist hydrogeological support if required.

5.11.3. Surface water and sediments

- For fresh water, ensure sample locations are upstream of the source of contamination.
- Consider the impacts of tidal flow, stratification, other sources of contaminants and the potential re-suspension of contaminants that have sorbed to sediments.
- Assess the weather before and at the time of sampling, and any potential impacts of the weather on the assessment.
- Collect and record physico-chemical parameters at all sampling locations. Compare sediment compositions.

Sampling of surface waters and sediments should be undertaken with reference to ANZG 2018 and Simpson and Batley 2016.

6. Hotspot detection

Hotspots are localised areas with significantly higher contaminant concentrations than other areas of a site or decision area.

Systematic sampling to detect hotspots of specified shapes and sizes is required to characterise or validate sites or decision areas. The sampling grids are placed at regularly spaced intervals, as discussed in Section 5.2. The grid size and pattern required for site characterisation depends on what is already known about the site and described in the CSM, and the shape and size of the target hotspots.

To determine the grid size, NEPC 2013, B2 says that:

> *Determining grid size/sampling density from mathematical formulae (for example, Appendix D of Standard AS 4482.1–2005) is not an acceptable approach without consideration of likely contaminant distribution and acceptable hotspot size.*

The number of sampling locations required for site characterisation is based on the following principles:

- the number of samples derived from the systematic sampling is adequate to indicate the true value of other critical parameters of a contaminant distribution, such as the average concentration and variability
- the spacing between sampling locations is determined according to the conceptual model, the phase of the investigation, acceptable levels of uncertainty and the requirements of the risk assessment (BSI 2013).

If the land to be sampled is intended for subdivision, the minimum hotspot size for investigation should be no larger than the size of the proposed or likely land parcels. While lot sizes depend on location and development type, an average lot size of between 400 m^2 and 500 m^2 is a reasonable assumption for urban areas.

This concept is, in part, derived from NEPC 2013, B2, which says:

> *If a site is to be subdivided, the size of the subdivided lots should be taken into account when determining the sampling density. While predictions may be made on a 'macro' scale, residents or owners may seek information about their own particular area of land and the risks associated with this land, especially if the potential contamination on the original site was uneven in distribution and type.*

Hotspots rarely have sharply defined boundaries. Contamination, strata and fill types are often heterogeneous pockets across sites due to site features and past uses. Accordingly, the use of systematic sampling grids should allow for appropriate location and mapping of the materials at the site, to provide representative data and determine the 'true value' of other critical parameters of a contaminant distribution, rather than finding distinctly definable hotspots.

For hotspot detection, samples collected from all sampling locations must be submitted for laboratory analysis.

For methods for determining the required grid size for circular and elliptical hotspots, see Appendix C: Determining sampling grids for hotspot detection.

The recommended number of sampling points required for site characterisation in Appendix C should not be considered as fixed: for irregularly shaped sites, more sampling points may be needed to detect hotspots of the calculated minimum size. As the number of sampling points required is in part based on the geometry of the site or decision area, the number of sampling points required depends on applying the specified grid size to the site or decision area. This method is illustrated in Figure 6.

Figure 6 **Placement of sampling grid and randomly selected sampling locations**
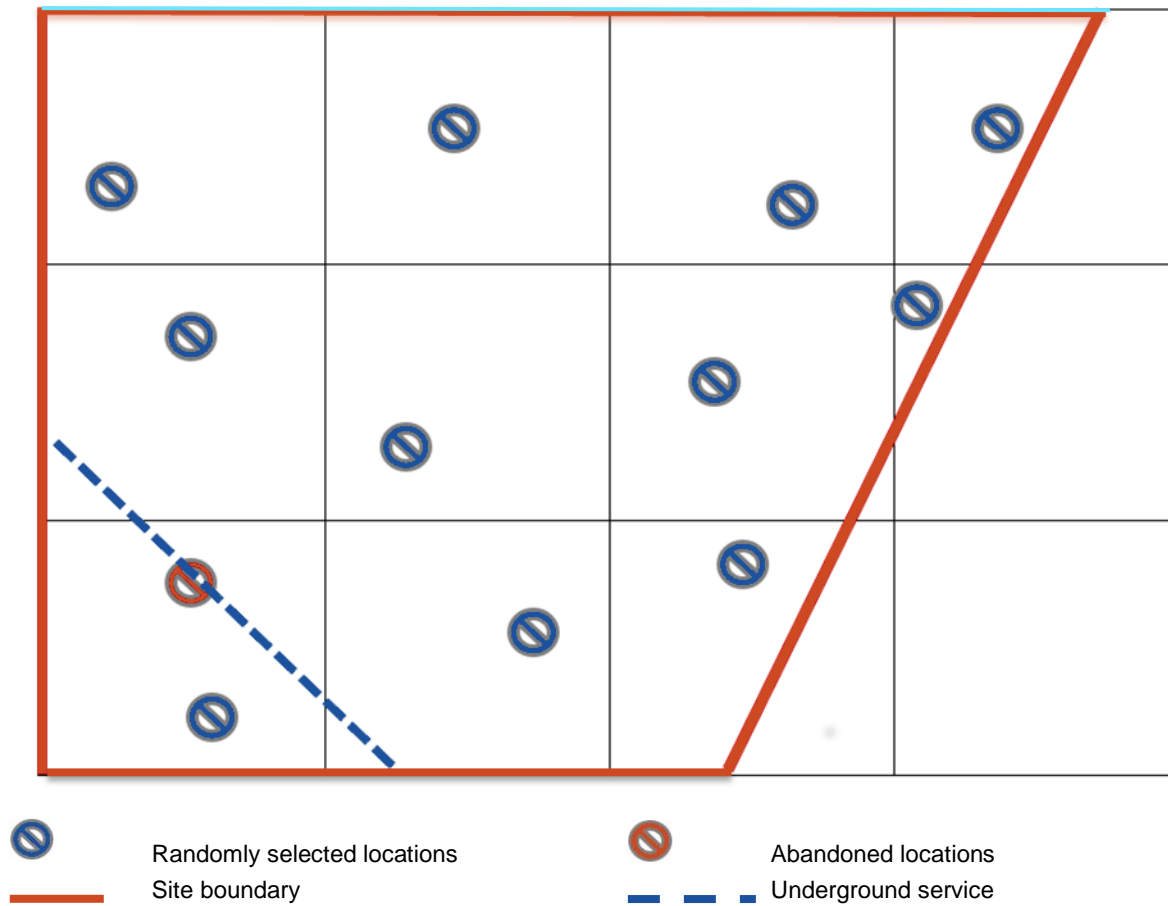
Source: Easterly Point Environmental Pty Ltd

When deciding whether to use a square grid pattern, consider the site characteristics and specific investigation objectives.

With systematic random designs, the randomness of the sampling points can be maintained by simply generating a new random coordinate set, if the design location is obscured. This is shown in Figure 6.

7. Number of samples required

The aim of environmental sampling is to collect sampling data that is appropriately representative of the population being sampled. **Representativeness** is a measure of the degree to which the sampling data accurately and precisely represents the characteristics of the population. The probability of achieving representative data is partly controlled by the number of samples.

This section provides methods for calculating the number of samples required to be representative of a population, by considering factors such as variance in the dataset and confidence levels. To determine the number of samples required for site characterisation or site validation as a function of the site area, see Section 5.2.

The number of samples required is defined by interacting factors including:

- the purpose of the sampling
- the sampling strategy selected
- the inherent variability of the target population
- the minimum effect size that needs to be determined
- the certainty required, including both the specified confidence level and the statistical power.

Representative analytical samples may have been collected via a systematic sampling regime, but it may be difficult to draw conclusions regarding the contamination status of the site based on the analytical results. For instance, there may be a large range of results, some results are greater than the assessment criterion (or criteria) or the 95% UCL is greater than the assessment criterion (or criteria). In these instances, more sampling **might** be justified, but the consultant needs to know how many samples are required.

The above factors should all be considered when determining the number of samples required to achieve an investigation's objectives.

The effect of some of these factors is illustrated in Figure 7, which shows the sample size needed for a one-sample t-test at a 95% confidence level and at various statistical powers (α = 0.05, β = 0.05, 0.1 and 0.2) (from USEPA 2002b, G–5S). See the glossary for the definitions of α and β. The number of samples required increases significantly as the effect size becomes a smaller fraction of the estimated value, and as the required confidence level and power increase.

Figure 7   Sample size at 80–95% confidence level, based on effect size as a fraction of the estimated value and power required
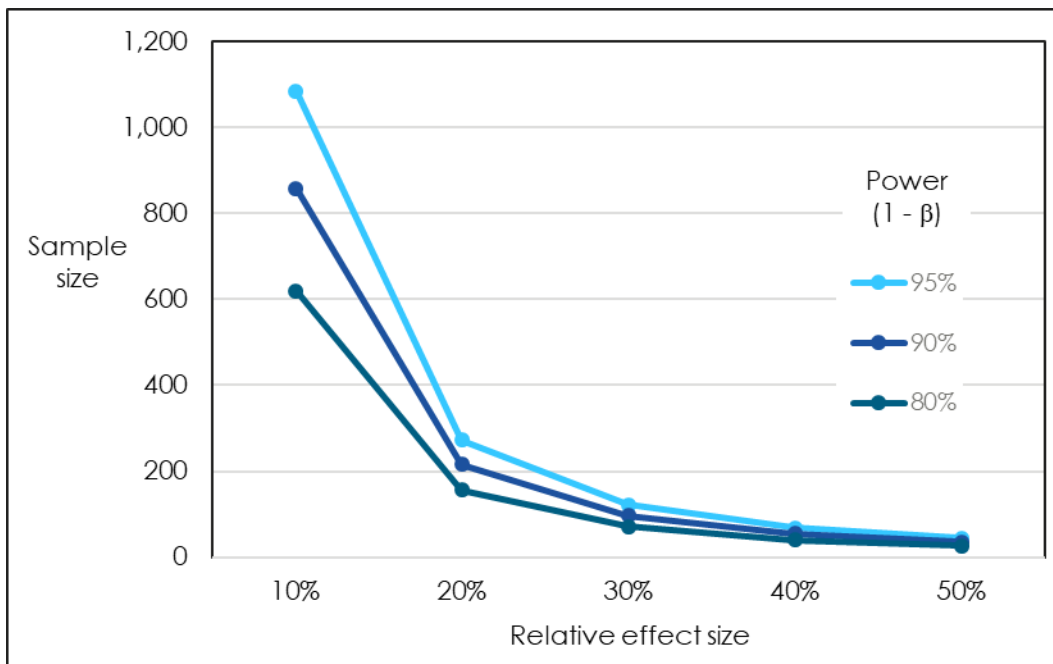
Source: Easterly Point Environmental Pty Ltd

Two statistical methods are provided for determining the number of samples (n) required: the CRV (combined risk value) method shown in Appendix E: Determining the number of samples by the CRV method and the MPE (maximum probable error) method shown in Appendix F: Determining the number of samples by the MPE method. Neither of these methods are based on the area of the site, but the calculation is performed using statistical parameters.

The CSM should be used or refined to show that one population is being considered for statistical treatment. If necessary, the consultant should attempt to stratify the site to identify distinct populations in the data.

The CRV method is in part determined by the **effect size**, which is defined as the magnitude of the difference between the populations or groups being studied. In the assessment of site contamination, the effect size typically measures the difference between, for example, the 95% UCL$\bar{\text{x}}$, and the criterion or action level. Many of the procedures used to determine n will provide a small n for large effect sizes, or a large n for small effect sizes. This includes the CRV method, which may provide an unrealistically small (<0.1) or large (>1,000) value for n based on the effect size. Accordingly, this method may not be suitable for determining if a site or decision area has been adequately characterised, or meets a specified criterion, unless other methods are used to confirm n. Both the MPE method and hotspot detection approach should also be used, where appropriate, as part of the multiple lines of evidence/weight of evidence approach.

Within the DQOs process, the phenomenon of the sample size increasing as the effect size becomes smaller is addressed through the use of the **grey region**. This is where the results are 'too close to call' (USEPA 2006b, G-4), and the consequences of making a decision error are considered to be relatively minor. USEPA 2000a, G-4HW describes the grey region as 'the range of possible parameter values near the action level where the cost of determining that the alternative condition is true outweighs the expected consequences of a decision error'.

In hypothesis testing, the width of the grey region is called the **minimum detectable difference Δ** (the uppercase Greek letter delta). It is determined by the parameter values for which the α and β probabilities are set and is the region for which decision errors are considered tolerable. In general,

the narrower the grey region, the greater the number of samples needed to determine whether to accept or reject the null hypothesis, H_0.

As H_0 indicates that the site is contaminated, the grey region represents the probability of a Type II or false acceptance decision error, and values within this region have a higher probability of being falsely accepted. When a $UCL\bar{x}$ is used for hypothesis testing, the probability of making a Type I or false rejection decision error is controlled; however, this approach does not control against making a Type II or false acceptance decision error.

Described more simply, H_0 means the site is contaminated and rejecting H_0 means the site is not contaminated.

If the decision rests on showing that the $UCL\bar{x}$ is less than the criterion, the number of samples required will depend on how close the arithmetic mean is to the criterion. The narrower the gap between the mean and the criterion, the more samples will be required to statistically demonstrate that the $UCL\bar{x}$ is less than the criterion.

See the Glossary for definitions of statistical terms used above, and Appendix A of Part 2 for a brief review of commonly used descriptive statistics.

7.1. Existing guidance

Appendix D: Summary of existing guidance for sample design summarises EPA-made and EPA-approved guidance on sample design, and other relevant guidelines.

When using the CRV and MPE methods to assist in determining the number of samples required to achieve the project objectives, take into account the media and/or contaminant types and incorporate it into the sampling design where relevant.

Neither the methods described below nor those referenced in Appendix D: Summary of existing guidance for sample design are to be considered minimum requirements. Rather, the method to be used needs to be chosen according to the situation-specific requirements of the investigation, and to be fully explained and documented, including any assumptions and limitations.

Spatial dependence is not considered in the CRV and MPE methods. These methods consider data that has already been collected probabilistically, for example, samples collected on a systematic grid for in situ sampling. Statistical parameters such as coefficient of variance (CRV method) or standard deviation (MPE method) are used to calculate the number of samples that would be needed to determine if a 95% UCL of a dataset is below a particular criterion.

7.2. Combined risk value method

The number of samples needed to show that the average concentration of a contaminant is below a defined criterion or action level can be determined using the CRV method. The CRV method can be used for a variety of media samples.

The determination is based on the principle of hypothesis testing, with the alpha (α) value for a Type I error, or false rejection of the hypothesis, and the beta (β) value for a Type II error, or false acceptance of the hypothesis, being used to determine the CRV. As the methodology is based on parametric methods, it assumes nearly-normal distribution and independent and unbiased sampling data.

The CRV method is used in hypotheses testing of the arithmetic mean to determine if n is sufficient. It is based on the specified values of α and β, and the effect size resulting from the difference between $\bar{x}$ and the specified criterion or action level. If the null hypothesis is not rejected, then the only potential decision error is false acceptance (β), and the CRV method can be used to determine if the error rate has been satisfied (USEPA 2006b, G-4). If n as determined by the CRV method is less than the number of analytical samples, then a Type II error may have been made. In such a case, the only way to maintain the selected probability is to increase the number of analytical samples.

Appendix E: Determining the number of samples by the CRV method shows how to use the CRV method to determine the number of samples required and gives a worked example.

7.3. Maximum probable error

When the objective of the sampling includes the estimation of the population arithmetic mean at a specified confidence level, the MPE method as described in Provost 1984 and Gilbert 1987 can be used. This method uses the margin of error (MoE), the standard deviation(s), and the t critical value, at a 95% confidence level or higher.

As MPE is based on parametric methods, it assumes nearly-normal distribution and independent and unbiased sampling data. The MPE equation ultimately reduces to n = n, that is, all other parameters cancel out. MPE cannot retrospectively demonstrate sufficient sampling, but provides a guide to an appropriate number of samples based on the variability of the data (standard deviation, s), and the required precision of the data (MoE). Once the standard deviation of the sample dataset is known, the desired MPE can be selected, and the number of samples required to achieve that MPE can be determined.

The MPE method can be used for all media, areas and stockpiles. However, it is insensitive to the area or volume of interest, and should be used in conjunction with other methods to confirm that a sufficient number of analytical samples has been collected and analysed.

Appendix F: Determining the number of samples by the MPE method shows how to use the MPE method to determine of the number of samples required and gives a worked example.

8. Abbreviations and glossary

8.1. Acronyms and abbreviations

95% UCL	95 percent upper confidence limit
ABC	Ambient background concentration
ACM	Asbestos containing material
ANZG	Australian and New Zealand water quality guidelines
ASC	Assessment of site contamination
AST	Above-ground storage tank
CECs	Contaminants of emerging concern
CLT	Central limit theorem
CLM	Contaminated land management
CoPC	Contaminants of potential concern
CRV	Combined risk value
CSM	Conceptual site model
CV	Coefficient of variation
DNAPLs	Dense non-aqueous phase liquids
DQIs	Data quality indicators
DQOs	Data quality objectives
DSI	Detailed site investigation
DUs	Decision units
EPA	Environment Protection Authority
HIL	Health-based investigation level
HSL	Health screening level
ISM	Incremental sampling methods
LNAPLs	Light non-aqueous phase liquids
LOR	Limits of reporting
Metals	Arsenic (As), cadmium (Cd), chromium (Cr), copper (Cu), lead (Pb), mercury (Hg), nickel (Ni) and zinc (Zn)
MoE	Margin of error
MPE	Maximum probable error
MQOs	Measurement quality objectives
NEPC	National Environment Protection Council
NEPM	National Environment Protection Measure
NHST	Null-hypothesis significance testing
NOW	New South Wales Office of Water
OEH	New South Wales Office of Environment and Heritage

PAHs	Polycyclic aromatic hydrocarbons
PFAS	Per- and poly-fluorinated alkyl substances
PFOS	Perfluorooctane sulfonate
PFOA	Perfluorooctanoic acid
PFHxS	Perfluorohexane sulfonate
PID	Photoionisation detector
PSH	Phase-separated hydrocarbon
PSI	Preliminary site investigation
PTFE	Polytetrafluoroethylene
QAPP	Quality assurance project plan
QA/QC	Quality assurance/quality control
Q-Q	Quantile–quantile
RAP	Remediation action plan
RSD	Relative standard deviation
SAQP	Sampling and analysis quality plan
SOPs	Standard operating procedures
STP	Sewage treatment plant
SWL	Standing water level
TOFA	Total organic fluorine assay
TOPA	Total oxidisable precursor assay
TRHs	Total recoverable hydrocarbons, including volatile C6–C10 fractions and semi- and non-volatile C11–C40 fractions
UCLs	Upper confidence limits
UCL$\bar{x}$	Upper confidence limits of means
UPSS	Underground petroleum storage system
USEPA	United States Environmental Protection Agency
UST	Underground storage tank
VOCs	Volatile organic compounds

8.2. Statistical notations

$1 - \alpha$	Confidence level
α	Type I error rate (see Glossary)
β	Type II error rate (see Glossary)
c	Criterion/action level
df	Degrees of freedom
exp	Exponential function
H_A	Alternative hypothesis
H_0	Null hypothesis

n	Number of samples or measurements in a sample (see sample definition)
θ	Scale parameter of the gamma distribution
σ	The population standard deviation, which is generally not known
σ^2	The population variance, which is generally not known
p-value	Probability value
Δ	Uppercase Greek letter delta, denoting the width of the grey region associated with hypothesis testing
s	The sample standard deviation, which is determined from the measurements taken
s^2	The sample variance, which is determined from the measurements taken
δ_0	Difference (delta) of zero
$t\alpha$	Critical value
t_0	Test statistic
μ	The population mean, which is generally not known
$UCL\bar{x}$	Upper confidence limit of arithmetic mean
$\bar{x}$	The sample mean, which is determined from the measurements taken
x_i	The ith measurement in the dataset

8.3. Glossary

95% UCL

The 95% upper confidence limit (UCL) is described in NEPC 2013, B1 Section 3.2.1 as follows: 'The 95% UCL of the arithmetic mean provides a 95% confidence level that the true population mean will be less than, or equal to, this value. The 95% UCL is a useful mechanism to account for uncertainty in whether the data set is large enough for the mean to provide a reliable measure of central tendency. Note that small data sets result in higher values for the 95% UCL.'

α risk

The probability, expressed as a decimal, of making a 'Type I error' when the hypothesis is tested statistically. A Type I error wrongly rejects a null hypothesis when in fact the null hypothesis is true. In this document, the null hypothesis always assumes that the site is 'contaminated' and thus the α risk refers to the probability of a site being validated as 'uncontaminated' when in fact it is 'contaminated'.

β risk

The probability, expressed as a decimal, of making a 'Type II error' when a hypothesis is tested statistically. A Type II error wrongly accepts a null hypothesis when in fact the null hypothesis is false. In this document, the null hypothesis always assumes that the site is 'contaminated' and thus the β risk refers to the probability that a site is concluded 'contaminated' when in fact it is 'uncontaminated'.

Acceptable limit

A threshold concentration value below which the level of contamination is regarded as acceptable. An acceptable limit can either be adopted from the appropriate guidelines or derived on a site-specific basis using risk assessment. Where site remediation is involved, acceptable limits are often referred to as 'clean-up standards' or 'remediation standards'.

Acceptance criteria

A statistical statement specifying how a contaminant distribution will be compared with an acceptable limit (see above definition) to determine whether a site should be evaluated as 'contaminated' or 'uncontaminated'. The concentrations of a contaminant can vary over orders of magnitude in a sampling area. All site assessments must state the appropriate acceptance criteria, as well as the appropriate acceptable limits.

Ambient air

External air environment, not including the air environment inside buildings or structures.

Arithmetic mean

The arithmetic mean is commonly referred to as the average and is used to describe the centre of the data distribution. It is obtained by adding all the values and dividing the result by the number of values.

Central tendency

The central or typical value for a probability distribution – it may be considered the average value in a set of data. It is generally described by the mode, median, or, more commonly, the mean, and describes where a sample distribution is centred.

Chi-squared distribution

A type of cumulative probability distribution that varies depending on the degrees of freedom (df). It is used to test relationships between categorical variables in the same population.

Coefficient of variation (CV)

CV is the measurement of the relative homogeneity of a distribution. Low CV values, for example, 0.5 or less, indicate homogenous contaminant distribution, while CVs with values more than 1–1.2 imply that the concentration distribution of a contaminant is heterogeneous and probably highly skewed to the right. A CV of more than 1.2 suggests that the data is lognormally distributed.

Composite sample

The bulking and thorough mixing of soil samples collected from more than one sampling location to form a single soil sample for chemical analyses.

Conceptual site model (CSM)

A CSM provides a three-dimensional overview of the contamination at sites and their surroundings, highlighting the sources, receptors and exposure pathways between the sources and receptors.

Confidence level

The probability, expressed as a percentage, that a statistical statement is correct. Confidence level is the opposite expression of 'risk' (see definitions of α and β risks). For the purpose of this document in which a risk that needs to be regulated, the confidence level is always equal to I - α.

Contaminated

For the purpose of this document and depending on the context, 'contaminated' can have slightly different meanings. If a site or a sampling area is evaluated as 'contaminated', it means that the site or the sampling area as a whole has not met the acceptance criteria (see definition of acceptance criteria). 'Contaminated' can also be used to describe a localised area or soil that has contaminant concentrations exceeding an acceptable limit (see definition of acceptable limit). Note: depending on what the acceptance criteria are, an entire site could be considered 'uncontaminated' even though a certain percentage of the site is expected to be 'contaminated'.

Data quality objectives (DQOs)

A systematic planning process used to define the type, quantity and quality of data needed to support decisions relating to the environmental condition of a site or a specific decision area.

Decision area

A specific area or medium on-site, or offsite about which data is being gathered so a decision can be made. For example, a decision can include part of a site, soil, a stockpile, soil gas, groundwater, surface waters or sediments.

Estimate

An estimate is a value that is inferred for a population based on data collected from a sample of units from that population. For example, the measured data from a sampling event used to calculate the sample mean ($\bar{x}$) is then used to estimate the population mean (μ).

Estimation

A technique that systematically adjusts the sample data to determine an estimated value for the population.

Geometric mean

This is similar to the arithmetic mean described above, in that it is also a measure of the central tendency of the distribution of a population or sample. It is sensible to calculate the geometric means only on populations or samples that contain positive values.

Grab samples

Samples collected from different locations that will not be composited but analysed individually.

Hotspot

A localised area where the level of contamination is noticeably greater than in surrounding areas.

Inter well

Comparison between two groundwater monitoring wells that are separated spatially.

Intra well

Comparison of measurements over time at one groundwater monitoring well.

Maximum

The maximum observed value in data, which generally provides a conservative estimate of the potential exposure risks. If the maximum is below the action level, the site should be suitable for the associated land use.

Median

The middle value of the distribution. Half the data values are less than the median and half are greater.

Minimum size effect

The acceptable magnitude of the difference between the populations or groups being studied.

Mode

The value that occurs most frequently. It is determined by counting the number of times each value occurs.

Modules

A series of discrete DQOs outputs, based on logical categories, that address selected components of a site investigation. Modules can be selected for contaminant types, media, decision areas or a combination of these.

Multiple lines of evidence

The process for evaluating and integrating information from different sources of data that uses best professional judgement to assess the consistency and plausibility of the conclusions which can be drawn. See weight of evidence.

Neyman–Pearson method

A method of statistical inference used to determine if a null hypothesis (H_0) should be rejected in favour of an alternative hypothesis (H_A), at a specified level of confidence.

Outlier

A data point that sits outside the expected range of the data. An outlier can have either a high or low value. Unless there is a demonstrable reason for rejecting it (such as coding error, sample contamination or equipment failure), an outlier needs to be retained within sample datasets. See Section 2.4 of Part 2 of these guidelines.

Parameters

Numerical measures of the characteristic of interest in the population being sampled. Typical parameters are the population mean (μ), variance (σ^2) and standard deviation (σ). Parameter values are usually unknown.

Percentiles and quartiles

These are descriptive values used to equally split a dataset into 100 parts. A percentile is the value that a given percentage of observations in a dataset is equal to or less than, for example, 80% of observations in a dataset are at or below the 80th percentile, while 20% are above.

Quartiles are commonly used to break the dataset up into four equal parts, providing an indication of the distribution and variance of the data.

> First quartile – the 0th percentile up to (and including) the 25th percentile
>
> Second quartile – from the 25th percentile up to (and including) the 50th percentile
>
> Third quartile – from the 50th percentile up to (and including) the 75th percentile
>
> Fourth quartile – from the 75th percentile up to (and including) the 100th percentile

Population

Any large collection of objects, things or individuals with some characteristics in common, that is being studied and for which information is sought. The population must be clearly and succinctly defined to allow effective sampling design and subsequent reporting.

The population can be further defined as the target population and the sampled population, and ideally these should be the same. The target population is the set of all units that comprise the items of interest, that is, the population about which a decision is required, and the sampled population is the part of the target population that is accessible and available for sampling. If the two diverge significantly, the target population should be redefined.

Probabilistic sampling

Probabilistic sampling occurs when each member of the population has a given probability (greater than zero and less than one) of being included in the sample. If the probability is the same for all population members the sample will be unbiased. Because inclusion in the sample is based on probability, subsequent samples will not necessarily include the same members.

Range

The range of a dataset measures the spread between the highest and lowest values in the dataset. Other measures (such as the standard deviation and the interquartile range) are required to provide an understanding of the distribution of the data.

Residual soil

The soil at a site that is not contaminated by industrial, commercial, or agricultural activities, consistent with the term 'ambient background concentration' (ABC) from NEPC 2013. Residual soil generally refers to soil that forms due to weathering or geomorphological processes, but can include reworked natural soils and historically imported material. Residual soils may have naturally occurring background levels of contaminants, contaminants that have been introduced from diffuse or non-point sources by general anthropogenic activity, and only low levels of contaminants attributed to industrial, commercial, or agricultural activities.

Sample

'Sample' has a number of meanings in the assessment of site contamination, including:

- as more broadly used in statistics, a representative group drawn from a population for description or measurement
- a physical amount of a material, for example, soil, water or air, or an aliquot taken for testing or chemical analysis
- a sampling point or sample location, being the location in plan at which a sample is collected, including description (e.g. geological logs) and field screening (e.g. PID, XRF).

Sample size

The number of samples or sampling points selected in a sampling program.

Sampling and analysis quality plan (SAQP)

Incorporates the CSM and the DQO outputs, to provide the context and justification of the selected sampling and analysis. The methods, procedures and quality control (QC) samples associated with the DQIs, including the frequency and MQOs, along with any associated contingencies, are also documented. The SAQP ensures that the data collected is representative and provides a robust basis for site assessment (NEPC 2013).

Sampling pattern

The locational pattern of sampling points within a sampling area.

Sampling point

The location at which a sample is collected.

Site characterisation

The assessment of the nature, level and extent of contamination. A typical site characterisation involves a preliminary site investigation (PSI), followed by a detailed site investigation (DSI), where warranted.

Site validation

The process of showing that a site is successfully remediated.

Standard deviation

Calculated by taking the square root of the variance (described below). It provides an indication of a population or sample data's typical deviation from its mean.

Statistic

Any summary number that describes the sample, such as an average or percentage. For example, the mean of a sample is described as x̄ (x-bar) and the standard deviation as s. When describing the population from which the sample is drawn, a summary number is called a parameter.

Statistical power

The probability of correctly determining a positive result, for example, a change or difference in the population, based on sample data. In this document, the statistical power is described as 1- β.

Sub-sample

A sample that will be combined with other sub-samples to form a composite for chemical analyses.

Systematic planning

A planning process based on a scientific method, which leads the project to unfold logically. Systematic planning includes established management and scientific elements. In the assessment of site contamination, it includes the application of the DQOs process and development of a CSM and SAQP.

Variable

A characteristic, number or quantity that is the subject of the inquiry. In the assessment of site contamination, it is usually continuous numerical variables that are being assessed, for example, the concentration of a contaminant in soil, soil gas or water. Discrete or discontinuous variables are at times considered, such as the number of fish in a waterbody. These are both **quantitative** variables in that they are derived by measurements.

Qualitative or categorical variables include ordinal or ranked variables and nominal variables. Ordinal variables are observations that take a value that can logically be ordered or ranked, such as first, second, third, whereas nominal observations take a value that cannot be organised in a logical sequence, such as presence or absence. Categorical variables are not commonly used in the assessment of site contamination and are not considered further.

Variance

The average squared distance of population or sample data points from the associated mean.

Weight of evidence/lines of evidence

'Weight of evidence' describes the process of collecting, analysing and evaluating a combination of different qualitative, semi-quantitative or quantitative lines of evidence to make an overall assessment of contamination.

Applying a weight of evidence process incorporates judgements about the quality, quantity, relevance and congruence of the data contained in the different lines of evidence (ANZG 2018).

A weight of evidence approach is where the consistency of data from more than one line of evidence is considered (NEPC 2013).

9. References

Australian and New Zealand Environment and Conservation Council (ANZECC) and Agriculture and Resource Management Council of Australia and New Zealand (ARMCANZ) 2000, *Australian and New Zealand guidelines for fresh and marine water quality*, paper no. 4, ANZECC and ARMCANZ, Canberra. Available at: www.waterquality.gov.au/anz-guidelines.

Australian and New Zealand Governments and Australian State and Territory Governments (ANZG) 2018, *Australian and New Zealand guidelines for fresh and marine water quality*, Water Quality Australia, Canberra ACT, www.waterquality.gov.au/anz-guidelines.

Beck PH, Mikov A and Curtis M 2004, Using spatial geostatistics to evaluate the risk contamination occurring between sampling locations, presented at Contaminated Site Remediation Conference, 16–18 September 2004, Adelaide.

British Standards Institution (BSI) 2013, *Investigation of potentially contaminated sites: code of practice*, BS 10175:2011+A1:2013, BSI Standards Limited.

Clements L, Palaia T & Davis J 2009, *Characterisation of sites impacted by petroleum hydrocarbons: national guideline document*, CRC Care Technical Report no. 11, CRC for Contamination Assessment and Remediation of the Environment, Adelaide.

CRC Care 2013, *Petroleum hydrocarbon vapour intrusion assessment: Australian guidance*, CRC CARE Technical Report no. 23, CRC for Contamination Assessment and Remediation of the Environment, Adelaide.

Crumbling DM 2001, *Applying the concept of effective data to environmental analyses for contaminated sites* (EPA 542-R-01-013), US Environmental Protection Agency, Washington DC, http://cluin.org/download/char/ effective_data.pdf.

Crumbling DM 2002, In search of representativeness: evolving the environmental data quality model, *Quality Assurance*, vol.9, pp.179–90, https://cluin.org/download/char/dataquality/dcrumbling.pdf.

Crumbling DM, Hayworth JS, Call BA, Davis WM, Howe R, Miller DS & Johnson R 2004, The maturing of the triad approach: avoiding misconceptions, *Remediation: The Journal of Environmental Cleanup Costs, Technologies & Techniques*, 14(4), pp 81–96.

Davis GB, Wright J & Patterson BM 2009, *Field assessment of vapours,* CRC CARE Technical Report no. 13, CRC for Contamination Assessment and Remediation of the Environment, Adelaide.

Department of Agriculture and Water Resources 2018*, Australian & New Zealand guidelines for fresh and marine water quality*, Department of Agriculture and Water Resources, Canberra, www.waterquality.gov.au/anz-guidelines.

Department of Environment and Conservation (DEC) 2004, New South Wales (NSW) *Australian river assessment system (AUSRIVAS) sampling and processing manual*.

Department of Environment and Conservation (DEC) 2005a, *Contaminated sites: Guidelines for assessing former orchards and market gardens*, DEC 2005/195, DECCW NSW, Sydney.

Department of Environment and Conservation (DEC) 2005b, *Information for the assessment of former gasworks sites*, DEC 2005/237, DECCW NSW, Sydney.

Department of Environment and Conservation (DEC) 2007, *Contaminated sites: Guidelines for the assessment and management of groundwater contamination*, DEC 2007/144, DEC NSW, Sydney.

Department of Environment, Climate Change and Water (DECCW) 2010, *Vapour intrusion: Technical practice note*, DECCW 2010/774, DECCW NSW, Sydney.

Department of Environment (DoE) Queensland 1998, *Draft guideline for the assessment & management of contaminated land in Queensland*, DoE, Brisbane.

Department of Environment and Science (DES) Queensland 2018, *Monitoring and sampling manual: Environmental protection (water) policy 2009* [sic], DES, Brisbane.

Department of Health and Ageing and enHealth Council 2012, *Environmental health risk assessment: guidelines for assessing human health risks from environmental hazards*, Department of Health and Ageing, Canberra.

Devore J & Farnum N 2005, *Applied statistics for engineers and scientists*, 2nd Edition, Brooks/Cole, Cengage Learning, Belmont CA.

Environment Protection Authority (EPA) 1995, *Contaminated sites: guidelines for the vertical mixing of soil on former broad-acre agricultural land*, EPA 2003/28, NSW EPA, Sydney.

Environment Protection Authority (EPA) 1997, *Contaminated sites: guidelines for assessing banana plantation sites*, EPA 97/37, NSW EPA, Sydney.

Environment Protection Authority (EPA) 2014a, *Resource Recovery Order under Part 9, Clause 93 of the Protection of the Environment Operations (Waste) Regulation 2014: The excavated natural material order 2014.*

Environment Protection Authority (EPA) 2014b, *Best practice note: Landfarming*, EPA 2014/0323, NSW EPA, Sydney.

Environment Protection Authority (EPA) 2014c, *Waste classification guidelines*, EPA 2014/0796, NSW EPA, Sydney.

Environment Protection Authority (EPA) 2015a, *Guidelines on the duty to report contamination under the Contaminated Land Management Act 1997*, EPA 2015/0164, NSW EPA, Sydney.

Environment Protection Authority (EPA) 2015b, *Technical note: light non-aqueous phase liquid assessment and remediation*, EPA 2015/0553, NSW EPA, Sydney.

Environment Protection Authority (EPA) 2017, *Contaminated land management: guidelines for the NSW site auditor scheme (3rd edition)*, EPA 2017/P0269, NSW EPA, Sydney.

Environment Protection Authority (EPA) 2018, *Guidelines on Resource Recovery Orders and Exemptions: for the land application of waste materials as fill*, EPA 2017/P0392, NSW EPA, Sydney.

Environment Protection Authority (EPA) 2019, *Standards for managing construction waste in NSW*, EPA 2019/P1542, NSW EPA, Sydney

Environment Protection Authority (EPA) 2020a, *Assessment and management of hazardous ground gases: contaminated land guidelines,* EPA 2019/P2047, NSW EPA, Parramatta.

Environment Protection Authority (EPA) 2020b, *Consultants reporting on contaminated land: contaminated land guidelines*, EPA 2020/P2233, NSW EPA, Parramatta.

Environment Protection Authority (EPA) 2020c, *Guidelines for implementing the Protection of the Environment Operations (Underground Petroleum Storage Systems) Regulation 2019*, EPA 2020/P2700, NSW EPA, Parramatta.

Environment Protection Authority (EPA) South Australia 2005, *Composite soil sampling in site contamination assessment and management*, Government of South Australia, Adelaide.

Environment Protection Authority (EPA) Victoria 2009, *Industrial waste resource guidelines: soil sampling*, IWRG702.

Ferguson CC 1992, The statistical basis for spatial sampling of contaminated land*, Ground Engineering*, vol. 25, no. 6, pp 34–38.

Gilbert RO 1987, *Statistical methods for environmental pollution monitoring*, John Wiley & Sons Inc, Brisbane.

Goovaerts P 1997, *Geostatistics for natural resources evaluation,* Oxford University Press, New York, p. 496.

Gray JM & Murphy BW 1999, *Parent material and soils: a guide to the influence of parent material on soil distribution in eastern Australia*, Technical Report No. 45, NSW Department of Land and Water Conservation, Sydney.

Hamon RE, Mclaughlin MJ, Gilkes RJ, Rate AW, Zarcinas B, Robertson A, Cozens G, Radford N & Bettenay L 2004, Geochemical indices allow estimation of heavy metal background concentrations in soils, *Global Biogeochemical Cycles*, vol. 18, GB1014.

Harr ME 1987, *Reliability-based design in engineering*, McGraw-Hill, New York.

Henley S 1981, *Nonparametric geostatistics*, Elsevier Applied Scientific Publishers, Barking UK.

HEPA 2018, *PFAS National environmental management plan version 2.0*, Heads of EPAs Australia and New Zealand. (Note: this document is updated frequently. For the latest guidance, see http://www.environment.gov.au/protection/chemicals-management/pfas.)

Interstate Technology and Regulatory Council (ITRC) 2007, *Vapor intrusion pathway: a practical guideline*, VI-1, ITRC Vapor Intrusion Team, Washington DC, USA, www.itrcweb.org/Documents/VI-1.pdf.

Interstate Technology and Regulatory Council (ITRC) 2012, *Incremental sampling methodology*, (ISM-1), ITRC, Washington DC, USA, www.itrcweb.org/Guidance/ListDocuments?topicID=11&subTopicID=16.

Interstate Technology and Regulatory Council (ITRC) 2013, *Groundwater statistics and monitoring compliance: statistical tools for the project life cycle*, GSMC-1, ITRC, Washington DC, USA, http://www.itrcweb.org/gsmc-1/.

Interstate Technology and Regulatory Council (ITRC) 2017, *Naming conventions and physical and chemical properties per- and polyfluoroalkyl substances (PFAS)*, ITRC, Washington DC.

Interstate Technology and Regulatory Council (ITRC) 2019 *Implementing advanced site characterization tools*, ITRC, Washington DC.

Isaaks EH & Srivastava RM 1989, *An introduction to applied geostatistics,* Oxford University Press, New York.

Krige DG 1981, *Lognormal-de wijsian geostatistics for ore evaluation*, South African Institute of Mining and Metallurgy, Johannesburg, South Africa.

Lock WH 1996, Composite sampling, in *National environmental health forum monographs, soil series no. 3*, South Australian Health Commission, Adelaide.

McDougall KW & Macoun TW 1996, *Guidelines for the assessment and clean up of cattle tick dip sites for residential purposes*, NSW Agricultural in conjunction with CMPS&F Environmental, Wollongbar NSW.

Naidu R, Jit J, Kennedy B & Arias V 2016, Emerging contaminant uncertainties and policy: The chicken or the egg conundrum, *Chemosphere*, vol. 154, pp 385–390.

National Environment Protection Council (NEPC) 2013, *National environment protection (assessment of site contamination) amendment measure 2013 (No. 1)*, Schedule A and Schedules B(1)–B(9), National Environment Protection Council, Canberra.

Nickerson RS 2000, Null hypothesis significance testing: A review of an old and continuing controversy, *Psychological Methods*, vol. 5, no. 2, pp 241–301.

Nielsen D & Wendroth O 2003, *Spatial and temporal statistics: sampling field soils and their vegetation*, Catena-Verlag, Reiskirchen, Germany, p. 614.New Jersey Department of Environmental Protection (NJDEP) 2005, *Vapor intrusion guidance.*

Northern Territory Environment Protection Authority (EPA) 2013, *Guidelines on conceptual site models*, NT EPA, Darwin.

Perezgonzalez JD 2015, Fisher, Neyman–Pearson or NHST? A tutorial for teaching data testing, *Frontiers in Psychology*, vol. 6, article 223, p. 1.

Provost LP 1984, Statistical methods in environmental sampling, in Schweitzer GE and Santolucito JA (eds), *Environmental sampling for hazardous wastes*, American Chemical Society, Washington DC.

Rayment G and Lyons D 2011, *Soil chemical methods - Australasia*, CSIRO Publishing, Melbourne.

Reinhart A 2015, *Statistics done wrong: the woefully complete guide*, No Starch Press, San Francisco CA.

Rendu JM 1981, *An introduction to geostatistical methods of mineral evaluation*, South African Institute of Mining and Metallurgy, Johannesburg, South Africa.

Royle AG 1980, *Geostatistics*, McGraw-Hill, New York.

Simpson S and Batley G (eds) 2016, *Sediment quality assessment: a practical guide*, CSIRO Publishing, Melbourne.

South Australian Health Commission (SAHC) 1995, *Guidelines for the composite sampling of soils*, SAHC, Adelaide.

Standards Australia 2005, *Australian Standard AS 4482.1-2005: Guide to the investigation and sampling of sites with potentially contaminated soil, part 1: non-volatile and semi-volatile compounds.*

The Centre for Research into the Built Environment (TCRBE) 1994, The Nottingham Trent University, *Department of the Environment contaminated land research report, sampling strategies for contaminated land*, CLR Report No 4.

US Environmental Protection Agency (USEPA) 1996, *Soil screening guidance: user's guide* (2nd edition), Attachment B, Soil Screening DQOs for Surface Soils and Subsurface Soils, EPA/540/R-96/018, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 1989, *Methods for evaluating the attainment of cleanup standards, volume 1: soils and solid media*, EPA 230/02-89-042, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 2000a, *Data quality objectives process for hazardous waste site investigations (QA/G-4HW)*, EPA/600/R-00/007, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 2000b, *Guidance on technical audits and related assessments for environmental data operations* (QA/G-7), EPA/600/R-99/080, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 2001, *Guidance on data quality indicators (QA/G-5i)*, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 2002a, *Guidance on environmental data verification and data validation (QA/G-8)*, EPA/240/R-02/004, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 2002b, *Guidance on choosing a sampling design for environmental data collection for use in developing a quality assurance project plan (QA/G-5S)*, EPA/240/R-02/005, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 2002c, *Guidance for quality assurance project plans (QA/G-5)*, EPA/240/R-02/009, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 2002d, *Overview of the EPA quality system for environmental data and technology*, EPA/240/R-02/003, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 2006a, *Data quality assessment: statistical methods for practitioners (QA/G-9S)*, EPA/240/B-06/003, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 2006b, *Guidance on systematic planning using the data quality objectives process (QA/G-4)*, EPA/240/B-06/001, Appendix: Derivation of sample size formula for testing mean of normal distribution versus an action level, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 2006c, *Data quality assessment: a reviewer's guide (QA/G-9R)*, EPA/240/B-06/002, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 2007, *Guidance for preparing standard operating procedures (SOPs) (QA/G-6),* EPA/600/B-07/001, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 2009, *Statistical analysis of groundwater monitoring data at RCRA facilities: unified guidance*, EPA 530/R-09-007, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 2014, *Test methods for evaluating solid waste: physical/chemical methods compendium (SW-846)*, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 2015, *ProUCL version 5.1.002: technical guide: statistical software for environmental applications for data sets with and without nondetect observations*, EPA/600/R-07/041, USEPA, Washington DC.

Webster R & Oliver MA 2001, *Geostatistics for environmental scientists*, Wiley, Chichester UK, p. 265.

Wendelberger J & Campbell K 1994, *Non-detect data in environmental investigations*, American Statistical Association, Toronto, Canada.

Western Australia Department of Health (WA DOH), 2009 *guidelines for the assessment, remediation and management of asbestos-contaminated sites in Western Australia*. WA DOH, Perth, WA

Wilson S, Card G & Haines S 2009, *Ground gas handbook*, Whittles Publishing, Dunbeath, UK.

Appendix A: DQOs and the environmental data life-cycle process

Environmental data life-cycle process

NEPC 2013, B2 recommends a systematic planning process be used to define the objectives of a site assessment, and a sampling plan be developed for collecting and evaluating representative data to achieve those objectives. It also states that the United States Environmental Protection Agency's (USEPA) seven-step Data Quality Objective (DQO) process is one example of a suitable systematic planning approach.

The DQOs process is one component of the USEPA's project-level quality system for collecting environmental data. This system includes various components which, taken together, form an environmental data life-cycle (EDLC) process for environmental assessments, and aims to produce 'defensible products and decisions'. Implicitly, partial or incomplete application of any individual components will result in data that are unlikely to achieve all the desired outcomes, that is, defensible products and decisions.

The components of the EDLC process, modified from USEPA 2002b, are shown in Figure 8 Environmental data life-cycle for assessment of site contamination (ASC) investigations and summarised below.

- **Systematic planning** – this identifies the expected outcome of the project, the technical goals, the cost and schedule, and the acceptance criteria for the final result before a project begins. The DQOs process includes developing or refining the Conceptual Site Model (CSM).

- **Sampling design** – this seeks to ensure that the data collection program collects appropriate and defensible data that accurately represents the problem being investigated. It is fundamental to data collection for scientifically based decision making.

- **Sampling, analysis and quality plan (SAQP)** – this documents the performance criteria and the project-specific plan for obtaining the type, quality, and quantity of data needed for a specific use. In addition to systematic planning and sampling design, inputs to the SAQP include:
  - Data Quality Indicators (DQIs) planning, to address the principal data quality attributes and the associated Measurement Quality Objectives (MQOs)
  - Standard Operating Procedures (SOPs) to document the procedures necessary to carry out routine or repetitive administrative and technical activities.

- **Conducting the study or investigation** – this is the implementation of the study or investigation, based on the preceding inputs. This can include technical assessments (project TQM audits), such as reviews to document the degree to which the procedures and processes specified in the SAQP are being implemented.

- **Data verification and validation** – this determines if data have been collected in accordance with the SAQP with respect to compliance, correctness, consistency, and completeness and evaluates the technical usability of the data with respect to the planned objectives or intention of the project, including in regard to the DQIs and MQOs.

- **Data analysis and interpretation** – this provides a scientific and statistical assessment to determine whether the data are of the right type, quality, and quantity to achieve the objectives of the project.

The EDLC process is iterative, and prior steps may need to be revisited based on the outcomes of later steps. Once the data have been verified and validated, as part of the data analysis and

interpretation, the design aspects should be reviewed to ensure any proposed statistical methods remain appropriate and that any assumptions made in the design phase are sustainable.

Figure 8   **Environmental data life-cycle for assessment of site contamination (ASC) investigations**
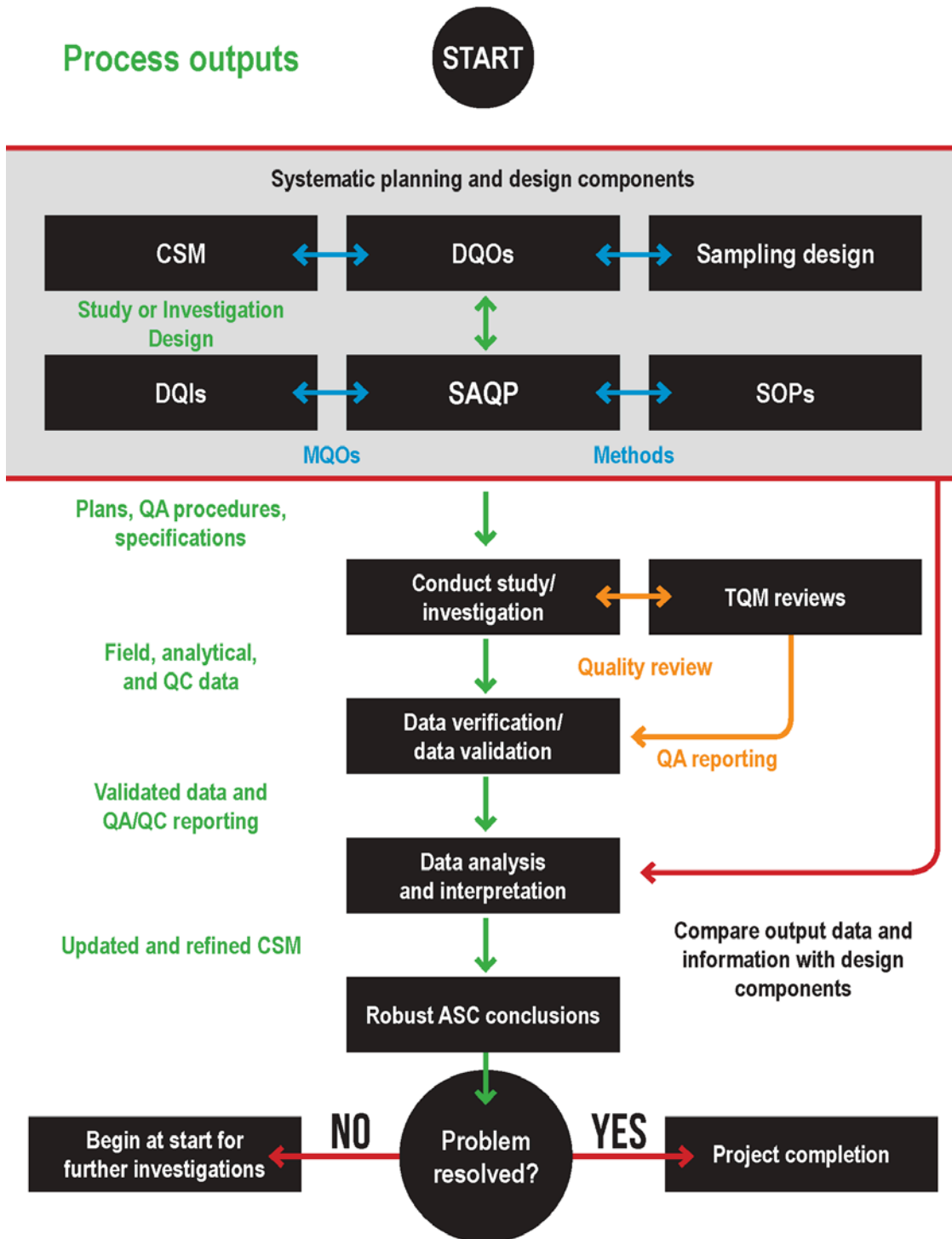
See text for definitions

Figure modified from USEPA G-4HW 2000a and www.epa.gov/quality/agency-wide-quality-system-documents#preview.

Project objectives

While project objectives are usually broad statements, such as 'to make the site suitable for its intended use' or 'to respond to a management order', these can be broken down into a series of distinct decisions to be made. The design components of the assessment of site contamination (ASC) generally address these decisions through investigations. Both the project objectives and the distinct decisions need to be considered and documented. The need to resolve numerous decisions arises as ASC projects generally address multiple contaminants, multiple media, and multiple potential receptors, often over many investigations, with many potential solutions for each decision.

Distinct decisions should be identified through developing and refining CSMs. This means the project objectives must be established; at the very least, the proposed land uses and whether the site's use is to be subject to long-term management should be specified. Changes to the project objectives, based on investigation results or modification of the proposed development or project, will generally change or modify the decisions required.

Distinct decisions to be answered should be as simple as possible to address a specific problem, as this allows the hypotheses to be directly stated and provides clarify in deciding if the problem has been resolved. For example, distinct decisions could be:

- Do contaminant concentrations in surface soils exceed the land use criteria?
- Does the soil stockpile meet the land use criteria?
- Is groundwater contaminated above the identified water quality objectives (WQOs) for the identified environmental values (EVs)?
- Is the surface water adjacent to the site being impacted by the on-site groundwater plume?

For complex problems, such as multiple contaminant types and a number of impacted media, more than one decision is generally required, or estimates of multiple parameters may need to be combined. These multiple decisions or estimates may combine or affect each other in resolving the problems. The DQOs process includes, in addition to CSMs, recommendations for the use of flow charts, logic diagrams and influence diagrams to illustrate, document and manage these problems.

For addressing multiple but specific technical questions, the use of modules is recommended, grouped by logical categories depending on the magnitude of the problem. Examples of categories include contaminant types, media, or decision areas, or a combination of these.

Decision problems and estimation problems

The seven-step DQOs process, as shown in Figure 9, is a method for systematic planning that includes options for the type of problem to be addressed, based on the intended use of the data to be collected. The two primary types of intended use are classified as **decision making** and **estimation**.

Decision making is defined as making a choice between two alternative conditions, for example, determining if site data are less than health investigation levels (HILs) or health screening levels (HSLs). USEPA 2006b, G-4 describes this process as:

> *This is where statistical methods help a decision maker structure the decision problem. The methodology of 'classical' Neyman–Pearson statistical hypothesis testing provides a framework for setting up a statistical hypothesis, designing a data collection program that will test that hypothesis, evaluating the resulting data, and drawing a conclusion about whether the evidence is sufficiently strong to reject or (by default) accept the*

hypothesis, given the uncertainties in the data and assumptions underlying the methodology. The DQO Process has been designed to support a statistical hypothesis testing approach to decision making.

Estimation is used when the objective of an investigation is to evaluate the magnitude of some environmental parameter or characteristic, noting that the resulting estimate may be used in further research, as an input to a model, or to support decision making. USEPA 2006b, G-4 notes that:

the defining characteristic of an estimation problem versus a decision-making problem is that the intended use of the estimate is not directly associated with a well-defined decision.

To illustrate the two types of problems, consider the project requirement to compare metal concentrations in soil at a site with the ambient background concentrations (ABCs). NEPC 2013, B5b states that the preferred method for estimating ABCs is by direct measurement at a clean reference site, with a soil type comparable to that of the site being examined. While no specific decision can be made as part of estimating the ABCs, some statistical rigour is desirable in estimating the metal concentrations. This is considered an estimation problem, with the number of samples required partly determined by the variance in the metals data, and partly by the required precision (margin of error).

Once suitable ABCs have been determined, the site data need to be compared to the ABCs. As a well-defined decision is required – whether the site concentrations are greater than, or not greater than, the ABCs – this is a decision problem. Project requirements could also include such things as comparing groundwater monitoring wells upgradient of a source to wells downgradient of a source, or comparing background surface water quality to the water quality of a release (or potentially contaminating discharge) into the surface waters. However, decision problems, such as comparing site data to specified levels, are the more common type of problem when assessing site contamination.

As shown in Figure 9, the DQOs process can broadly be classified as 'planning and input aspects' in Steps 1 to 4, and 'design aspects' in Steps 5 to 7. This is an iterative process, and revisiting earlier steps is often required. NEPC 2013, B2 and USEPA 2006b, G-4 should be referred to for specific details of the planning and input aspects, with the design aspects discussed below.

Figure 9   Overview of the USEPA DQOs process
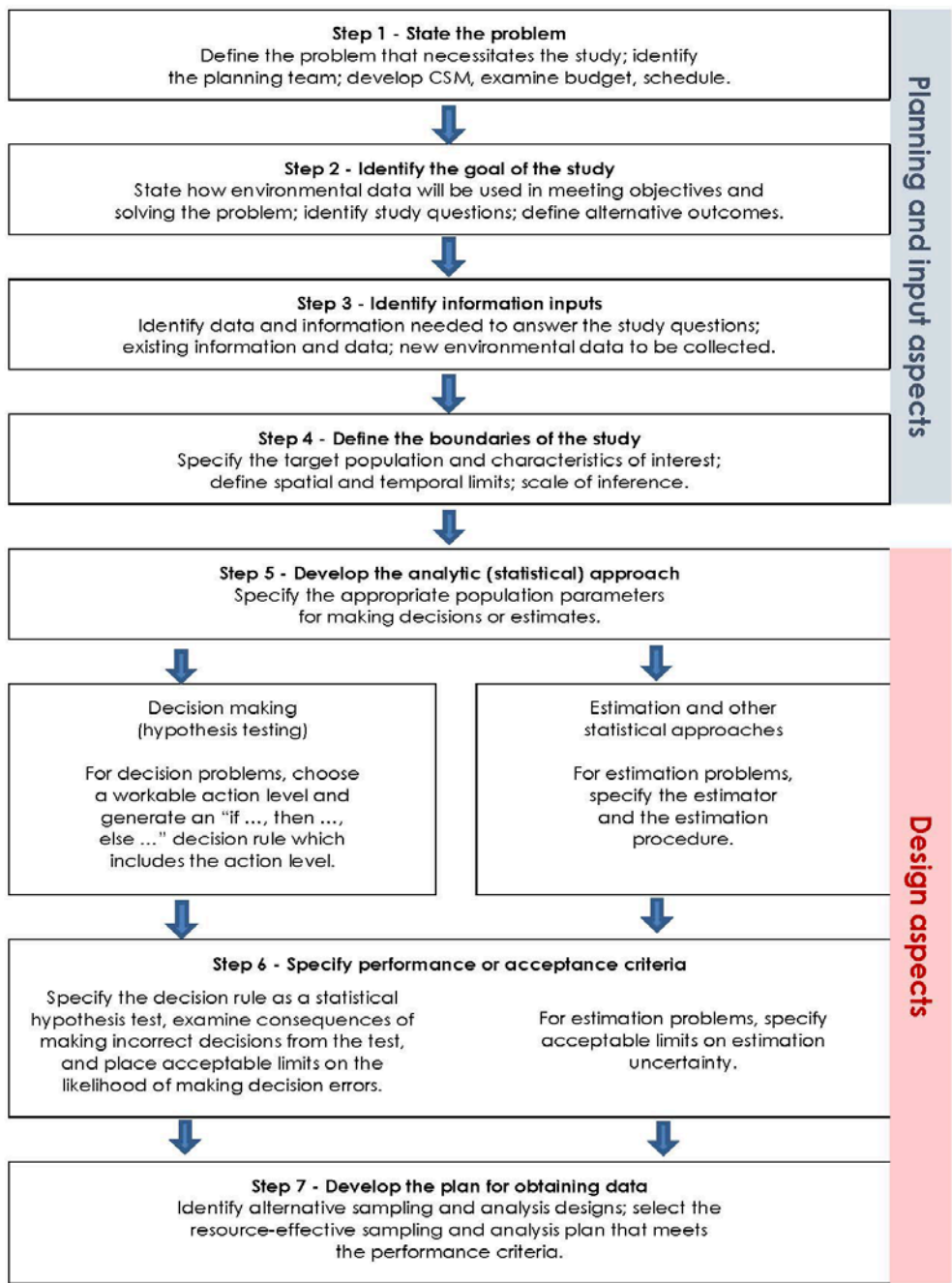
Source: Easterly Point Environmental Pty Ltd

Modified from USEPA (2006b, G-4)

Step 5 – Develop the analytic approach

Although at times confused with laboratory analysis of environmental media, Step 5 relates to 'developing an analytic approach that will guide how you analyse the study results and draw conclusions from the data' (USEPA 2000a, G-4HW), that is, it is in regard to analysing the primarily analytical chemistry data. USEPA 2000a, G-4HW gives the title of Step 5 as 'Develop a decision rule'. This title relates to specifying the statistical parameter that characterises the population of interest, such as mean, median, maximum, or proportion. G-4HW notes that the term 'statistical parameter' refers to the key characteristics of the population of interest, and that by definition, it is unknown and can only be estimated by measuring a similar characteristic from a sample of the population of interest.

The activities to be undertaken as part of Step 5 for decision problems are:

- Specify the relevant population parameter to make inferences about the target population (e.g. mean, 95% upper confidence limits (UCL) of the arithmetic mean, median or percentile).
- Choose an action level using the information identified in Step 3 (see Figure 9) that sets the boundary between one outcome of the decision process and an alternative, and verify that there are sampling and analysis methods that have detection limits below the action level.
- Construct the theoretical 'if ..., then ..., else ...' decision rule by combining the true value of the selected population parameter, the action level, the scale of decision making – see Step 4, Figure 9, and the alternative actions – see Step 2, Figure 9.
- For decision problems, the outputs for this step are:
  - o identification of the population parameters most relevant for making inferences and conclusions about the target population
  - o the 'if ..., then ..., else ...' theoretical decision rule based on a chosen action level.

For estimation problems, Step 5 involves specifying the estimator by combining the selected population parameter, for example, mean, with the scale of the estimation and other population boundaries from Step 4 (see Figure 9), then applying the estimation procedure, for example, 95% confidence interval.

Step 6 – Specify performance or acceptance criteria

Step 6 of the DQOs process establishes quantitative criteria known as performance or acceptance criteria, or data quality objectives (DQOs). The DQOs vary depending on the type of problem being addressed:

- for decision problems, the DQOs are typically tolerable limits on the probability or chance (risk) of the collected data leading to making an erroneous decision (e.g. confidence levels)
- for estimation problems, the DQOs are typically an acceptable uncertainty, for example, the width of an uncertainty band or interval, associated with a point estimate at a desired level of statistical confidence (e.g. confidence intervals).

USEPA 2006b, G-4 notes that performance criteria represent:

> *the full set of specifications that are needed to design a data or information collection effort such that, when implemented, generate newly-collected data that are of sufficient quality and quantity to address the project's goals.*

Acceptance criteria are:

> *specifications intended to evaluate the adequacy of one or more existing sources of information or data as being acceptable to support the project's intended use.*

Accordingly, the DQOs process should be used to generate **performance criteria** for new environmental data and **acceptance criteria** for existing information and data. Where existing data and information do not meet the acceptance criteria, they may need to be classified as estimates,

and new information and data may need to be obtained, subject to the specified performance criteria.

Step 7 – Develop the plan for obtaining data

Step 7 of the DQOs is to develop a resource-effective, field investigation sampling and analysis design to generate data that satisfy the decision performance criteria in Step 6 and the requirements specified in the preceding steps of the DQOs. It is usual to iterate between Steps 6 and 7 when assessing and refining the design parameters against the project objectives and constraints. The output of Step 7 is the sampling and analysis design that is documented in the SAQP.

For most field investigations, a probabilistic sampling approach is necessary to provide a scientific basis for extrapolating the results from samples to the entire site or decision area. USEPA 2000a states that by:

> *combining an effective probabilistic data collection design with a statistical hypothesis test, the decision maker will be able to optimize resources such as funding, personnel, and time while still meeting DQOs.*

For common probabilistic designs, information regarding the expected variability of the contaminants is necessary, as determining a minimum sample size relies on an estimate of total variability in the data to be collected (USEPA 2006b, G-4). Such estimates may be determined from existing data on the site or from similar sites. If no existing data are available, limited field investigations may have to be undertaken to determine a preliminary estimate of variability.

Information derived from the systematic planning that is used as input to the sampling and analysis design process includes:

- a description of the target population and the spatial/temporal boundaries of the study (DQO step 4)

- the preliminary estimation of variance of the target population (DQO step 4)

- the purpose of the data collection – hypothesis testing, estimating a parameter with a level of confidence, detecting hotspots or some combination (DQO step 5)

- the statistical parameter of interest, such as the mean, the 95% UCL of the arithmetic mean, the median, percentile, trend or slope (DQO step 5)

- limits on decision errors and precision, in the form of false acceptance and false rejection error rates and/or the overall precision specifications (DQO step 6).

Step 7 includes developing alternative data collection designs to assess which design best limits the total study error to tolerable levels to satisfy the decision performance criteria. To generate alternative designs, aspects to be considered include:

- type of samples collected

- sampling design

- sample selection technique

- number of samples

- spatial/temporal locations of samples

- field sampling or analytical methods used

- number of analyses per sample

- number of replicate analyses performed on samples.

USEPA 2000a, G-4HW states that two mathematical expressions are:

necessary for optimizing each data collection design alternative in relation to the decision performance criteria. First, a tentative method for analysing the resulting data (e.g. a student's t-test or a tolerance interval) should be specified, along with any available sample size formulas corresponding to the proposed method. This information will be used to solve for the minimum sample size that satisfies the decision maker's limits on decision errors. Second, a cost function that relates the total number of samples to the costs of sampling and analysis should be developed. This information will be used to compare the cost-effectiveness of different sampling designs.

Whereas statistical design generally addresses the collection of analytical samples, substantial information is generated by field samples and their associated observations, descriptions and field tests. USEPA 2000a, G-4HW notes that designs that 'balance the number of field samples with the number of laboratory analyses should be considered'. While field samples do not influence the number of samples determined by mathematical equations, which is a similar problem with non-detects, they need to be included in the robust CSMs, as they add to the weight of evidence and provide additional lines of evidence.

Weight of evidence is described in ANZG 2018 as 'the process to collect, analyse and evaluate a combination of different qualitative, semi-quantitative and quantitative lines of evidence to make an overall assessment of contamination. Applying a weight of evidence process incorporates judgements about the quality, quantity, relevance and congruence of the data contained in the different lines of evidence'.

It is also recommended that a sensitivity analysis be performed on the alternative designs, to determine if changes to design assumptions significantly affect the design's ability to achieve the expected decision error limits, and the associated impacts on costs or resources required. For example, if contaminant variability is higher than estimated, will the proposed number of samples meet the performance criteria or will more samples be required, leading to higher cost?

Once a data collection design has been selected, the design parameters and key assumptions must be documented so the collected data can be analysed and interpreted to determine if the data are of the type, quality and quantity required to achieve the project objectives. In the USEPA's environmental data life-cycle process, this occurs during the assessment stage as part of the data analysis and interpretation.

While the choice of the sampling and analysis design will have an impact on the DQIs, and the DQIs should be considered as part of Step 7, the DQOs do not specifically address the DQIs or their acceptance criteria. In much the same way as Step 5 of the DQOs is conducted under the assumption that one has 'access to perfect information on unlimited data' (USEPA 2006b, G-4), the DQOs process assumes that all the data collected is usable, at least until Step 7 and the subsequent development of the SAQP, as discussed below.

Data quality indicators and measurement quality objectives

Data quality is a measure of the degree of acceptability or usability of sampling data for a particular purpose. It relates to both sampling errors and measurement errors. As USEPA 2006b G-4 notes, sampling error is generally much larger than measurement error and consequently needs a larger proportion of resources to control.

Figure 10 shows an example of how total study error can be broken down into components that are associated with the various activities as part of environmental sampling and analysis. While interrelated, the activities associated with sampling error are predominately addressed through the DQOs process and sampling design, and the activities associated with measurement error are predominately addressed through the DQIs, MQOs and SOPs. The magnitude of total study error should be controlled by **generating an appropriate sampling design and choosing suitably accurate measurement techniques.**

In regard to measurement errors, certain qualitative and quantitative characteristics of the collected data, that is, the data quality attributes, can be defined and measured. DQIs are the quantitative

and qualitative measures, or indicators, of the principal data quality attributes. The principal data quality attributes are precision, accuracy[3], representativeness, comparability, completeness, and sensitivity (PARCCS), with precision, accuracy/bias and sensitivity being defined and measured in quantitative terms, and representativeness, comparability and completeness having more qualitative definitions. MQOs are the acceptance criteria or goals for the data quality attributes, and include such things as relative percentage differences (RPDs) and percentage recoveries of sample spikes.

USEPA (2001, G-5i) states that DQOs are qualitative and quantitative study objectives for the collection of environmental data, and that historically:

> *DQIs sometimes have been incorrectly equated with DQOs, which are specifications for decision making.*

DQIs are not the focus of this guidance but they are important inputs to the sampling design process, as they indicate whether the resulting data are expected to meet the DQOs, and the process of establishing MQOs, the goals set for the DQIs, is an integral part of designing the study.

After collecting the data, its adequacy or usability should be determined as part of the data-verification and data-validation component of the data life-cycle process. Use a weight of evidence/multiple lines of evidence approach and take into account both the project specific requirements and the stage of the data collection event. Data quality requirements for final data (characterisation or validation), will generally be more stringent than for preliminary data.

Figure 10   Total study error by components
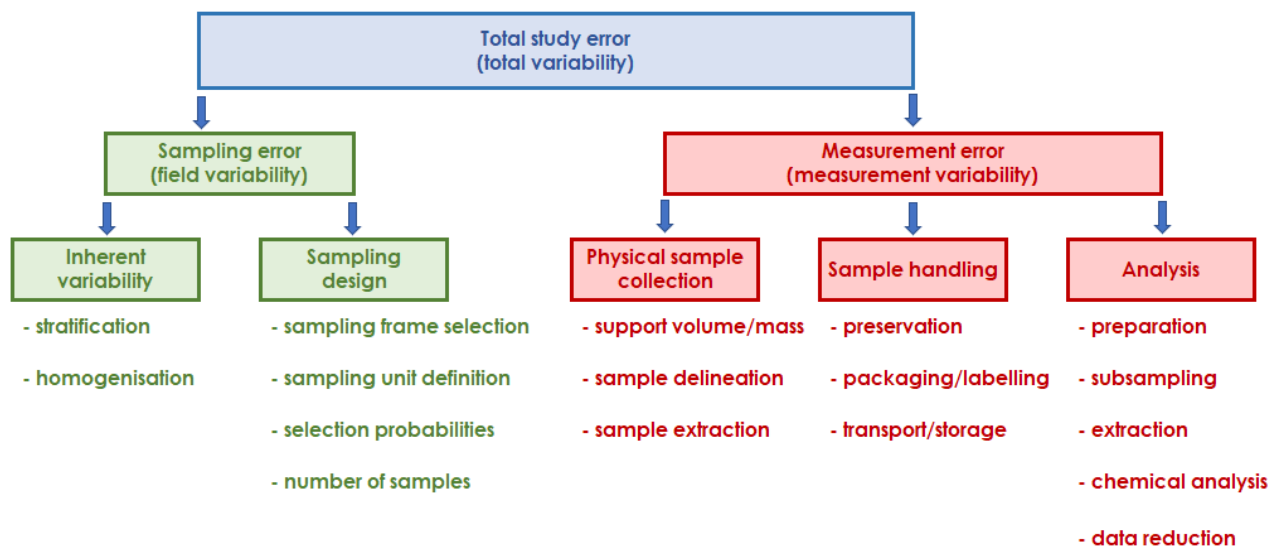
Green predominately relates to DQOs and red to DQIs, MQOs and SOPs
From USEPA 2006b, G-4

USEPA (2001, G5i) notes that:

> *the highest interest is in whether the data set will support a decision with the desired degree of certainty. It is important to consider the performance and representativeness*

---

[3] USEPA 2001,G-5i describes that the 'A' in PARCCS refers to accuracy instead of bias, and that this substitution of 'A' for 'B' occurs because PARCCS is a historically recognised and familiar acronym, and some analysts believe accuracy and bias are synonymous, although accuracy is actually comprised of random error (precision) and systematic error (bias).

of the measurement effort prior to reaching conclusions regarding data adequacy; however, at this point it is less critical to determine if each and every goal set for given DQIs (i.e. the MQOs) was achieved.

*If adequate sensitivity was achieved, and bias is 'under control,' **the key issues revolve around whether an adequate number of samples was obtained, given the observed measurement, spatial and temporal variability, and given the actual magnitude of the measurements made (relative to levels of concern)**. If a data collection effort fails to generate adequate data, then interest in DQIs is heightened.*

In this context, an adequate 'number of samples' must explicitly include an evaluation of the sample representativeness. **If the samples cannot be shown to be representative of the condition of the site or decision area, in the context of the decision to be made, evaluation of the measurement quality in isolation cannot demonstrate if the data are of a suitable quality to support the required decision** (Crumbling 2001).

Clarification of 'quality'

NEPC 2013 states that the DQOs process is used to define 'the type, quantity and quality of data needed'. However, within the DQOs process, 'quality' means the quality of the estimates derived from the data or refers to a desired level of quality such as the statistical precision of the data. As total study error is controlled by both sampling errors and measurement errors, both need to be understood and controlled to ensure defensible decisions.

At times, similar concepts have different practical applications, for example, representativeness and precision need to be controlled for both types of error. Representativeness addresses the extent to which measurements actually reflect the sampling unit from which they were taken (measurement error), as well as the degree to which samples actually represent the target population (sampling error). Therefore, the sampling error component of representativeness is addressed by properly specifying the number and location of samples within the study design. USEPA 2015, in discussing hypotheses testing approaches, highlights that 'good quality data' relates to representative data.

USEPA 2000a points out that the DQOs process represents an evolution from concerns about the quality of data to concerns about the quality of the decisions that will be made from the data. It notes that data quality:

as a concept, is meaningful only when it relates to the intended use of the data. Data quality does not exist without some frame of reference; one must know the context in which the data will be used in order to establish a yardstick for judging whether or not the data set is adequate.

The use of the term 'quality' by the USEPA is illustrated by its definition of data quality assessment (DQA), described as data analysis and interpretation in the EDLC process as shown in Figure 8 Environmental data life-cycle for assessment of site contamination (ASC) investigations. The five steps of statistical DQA are described by USEPA 2006c, G-9R as:

- review the project objectives and sampling design – undertaken by reviewing the systematic planning objectives to ensure the context of the investigation is understood. This review also allows the quality of the data to be assessed in terms of addressing the objective of the investigation as well as its quality for use

- conduct a preliminary data review – this includes an assessment of quality assurance (QA) reports to identify any anomalies – the data should also be assessed for its distribution and patterns, and to identify any potential outliers

- select the statistical method – this will be guided by the previous two steps; the choice of statistical method will be based on the objectives of the investigation and the dataset

- verify the assumptions of the statistical method to ensure that any assumptions made are justified, for example, if the dataset is highly skewed, some statistical methods may lead to biased conclusions and under-estimate key parameters

- draw conclusions from the data findings in line with the objectives of the investigation – the conclusions should include an assessment of the sampling design and whether or not it can be used in other scenarios; conclusions should be documented and coherently justified so all stakeholders can understand how they were reached.

The DQOs process, as designed and implemented by the USEPA, is a component of a multi-stage project life cycle that primarily addresses the sampling design and statistical aspects of a proposed environmental evaluation. The DQA component is a stage after the implementation of the data collection: it addresses the statistical rigour of the investigation and the achievement of the project objectives, and relies on both the data collected as part of the study or investigation, and the DQOs developed during the planning phase of the study or investigation, as shown in Figure 8 Environmental data life-cycle for assessment of site contamination (ASC) investigations.

Project level planning

At a broader, 'whole of project' level, assessment of site contamination projects also requires systematic planning, to ensure an appropriate level of project planning and documentation to manage the overall project. Projects often consist of separate investigations and usually involve sequential steps of assessment and management, from the preliminary investigation through to remediation, and ultimately validation and completion.

The systematic planning process adopted should identify the objectives of the site investigations and establish the types of information needed to make the necessary various environmental decisions. The DQOs process is often used in this context. For example, NEPC 2013, B2 states that the 'DQO process is applicable at both the project level (for example, is the site suitable for development?) and at the investigation level'.

As the DQOs process was designed for addressing multiple but specific technical questions, the later specific design steps of the DQOs do not generally apply to the broadscale, project level planning. The DQOs process can be applied in various ways at different stages of the project. USEPA 2000a points out that:

> during early site assessment phases, where investigators generally examine existing site information and conduct site reconnaissance, planning teams can benefit from the qualitative DQO steps, but may have to allow for a more liberal interpretation of the quantitative steps.

Table 6 gives the recommended approach for applying the DQOs process across all levels of site contamination assessment projects.

Table 6   Recommendations for implementation of DQOs

Project requirement	Applicable DQOs step
Project level	All steps, with only generalised information in steps 5, 6 and 7. Project level SAQP can also be developed, including standard operating procedures (SOPs) for sample collection and handling and well installation, etc.
Individual investigations	All steps, with steps 5, 6 and 7 fully addressed for simpler investigations. Document in an investigation-level SAQP or a project-level SAQP with specific investigation requirements added. A site-specific CSM should be developed and then refined at each investigation level.
Complex investigations with multi-contaminant types, media, or site histories, including risk assessment	Address project-level requirements at the project level and use modules to address specific technical questions.

investigations or offsite investigations of surface waters, sediments or biota	Document in either a project-level or investigation-level SAQP, with specific module requirements added.

Appendix B: Data-quality objectives: worked example

This worked example is a decision problem, based on the following hypothetical scenario.

A site operated as a sheepskin processing facility between the 1950s and the early 1970s. It had a combined storage shed/office/amenities building and a workshop that incorporated asbestos-containing cement sheeting. The sheepskins were dried on racks housed in timber and corrugated-iron sheds, and also directly on the ground.

Arsenic (As) was used as a biocide to treat the skins, and a 500-litre vat for mixing and storing it was in one of the drying sheds. The As was in powder form and mixed with water on-site, with the treatment solution applied using a network of irrigation pipes below the ceiling. The spent As solution was discharged to the land surface.

Anecdotal information also suggests some fuel was stored on the site, although it has not been confirmed if this was above ground or below ground. It is likely that other chemicals and fuels such as paints, solvents and greases were used in the workshop area. It is not known if a significant amount of waste was buried on-site, although some wastes have been dumped, for example, empty 200-litre drums. Several areas of disturbed natural soils, and some building and demolition wastes beneath former structures, have been identified.

Many investigations have been conducted in the area of the old processing facility, located at the crest of a small hill: they have included judgmental and systematic soil sampling, and judgmental groundwater sampling. These have found no organic contaminants or asbestos fibres but have identified a number of soil locations where elevated As occurs, as well as some copper (Cu) and zinc (Zn), assumed to be related to galvanised materials in the sheds, and some lead (Pb), assumed to be related to Pb paint flakes. No groundwater impacts were identified. The old processing facility has been considered to be sufficiently characterised to not require further investigations.

An area adjacent to the former processing centre is approximately 80 m x 160 m and has been potentially contaminated by surface overflow of process water. Potential contamination extends to between 2 and 3 m in depth. This area requires characterisation to determine the land-use suitability or the appropriate waste classification. The intended land use is residential with accessible soil (HIL-A), and it is proposed to develop the decision area into 400 m^2 residential blocks, although the specific lot layout has not yet been determined.

The DQOs process output for the proposed investigation for this portion of the site is shown in Table 7.

Table 7   DQOs process steps and their outputs

No.	DQO process step	Outputs of DQOs process step
1	**State the problem** – assemble an effective planning team, describe the problem and examine the resources for investigating the problem.	–
1.1	Write a brief summary of the contamination problem.	An area of land of approximately 80 m x 160 m and between a depth of 2 m and 3 m from the surface requires characterisation to determine the land use suitability or the appropriate waste classification.
1.2	Identify members of the planning team.	Landowner/developer, planning consultant and site contamination assessment consultant.
1.3	Develop/refine the conceptual site model (CSM), including a summary of the exposure scenarios.	Contaminants – potentially metals (As, Cu, Pb and Zn), organics (TRHs, BTEX, PAHs and OCPs/OPPs), and asbestos fibres/fragments. Sources – buried building and demolition wastes, former chemical wastes and drums, and overflow of waste process water. Receptors – site maintenance workers and trespassers, as the site is fenced and secured. If developed, site workers (surface and sub-surface), residents and visitors (adults and children). Pathways – dermal contact, inhalation of dust and ingestion have been identified as the pathways of concern. Further assessment of groundwater and/or soil gas will be considered based on the findings of this investigation.
1.4	Specify the available resources and constraints, such as relevant deadlines for the study, budget, availability of personnel and schedule.	The site contamination assessment consultant has the available capacity to conduct the investigation using appropriate subcontractors (excavator and laboratory). While the developer seeks close out of the issue within the next three months, there are no practical constraints, as the land is identified as high-value and a sufficient budget is available. Additional investigations, or remediation will be conducted as required.
2	**Identify the goals of the study** – identify the principal study question(s) and potential alternative actions (with implications), and combine these to make statements on the decision problem.	–
2.1	Identify the principal study question(s).	Is the material suitable for a residential land use based on contaminant levels and aesthetic concerns? If not, what disposal options are available, i.e. what is the waste classification?

No.	DQO process step	Outputs of DQOs process step
2.2	Identify the alternative outcomes or actions that could result from resolution of the principal study question(s).	The alternative outcomes will be: • the material is suitable for residential land use (HIL-A) or • the material is not suitable for the proposed land use and needs to be partially or fully removed from the site to allow for development.
2.3	For decision problems, combine the principal study questions and the alternative actions into decision statements.	If the contamination status of the material is acceptable, the material can remain on-site. If the contamination status of the material is unacceptable, consider the remediation hierarchy.
3	**Identify information inputs** – identify the information needed to formulate and investigate the problem and confirm that appropriate sampling and analytical methods are available.	–
3.1	Identify the information that will be required to resolve the decision statements/estimation, including existing information and new environmental data, and identify the sources for each item of information required.	Soil data collected as part of this investigation, including field samples and analytical samples. No previous investigation of the area has been conducted, although information from the investigation of the adjacent facility will inform this investigation.
3.2	Identify the information needed to establish the action level.	Investigation criteria will be sourced from: • NEPC 2013, Schedule B1, HILs for residential with accessible soil • NSW EPA 2014c, Waste Classification Guidelines.
3.3	Confirm that appropriate sampling and analytical methods exist to provide the necessary data.	Sampling and analytical methods will be consistent with existing guidance, including NEPC 2013, B2 and B3. Analytical laboratories will be NATA accredited and/or subject to proficiency testing and use analytical methods based on NEPC, USEPA and APHA methods.
4	**Define the boundaries of the study** – define the target population and the spatial and temporal boundaries associated with the population; examine any practical constraints to collecting data, and factors that affect the selection of the unit that	_

No.	DQO process step	Outputs of DQOs process step
	defines the scale of sampling and the scale of decision making or estimation.	
4.1	Define the target population of interest and its relevant spatial boundaries.	The area is approximately 80 m x 160 m and is between a depth of 2 m and 3 m from the surface. The decision area is approximately 12,800 m^2 and contains an estimated 32,000 m^3 of material. The natural soil is silty to sandy clay with frequent weathered parent material (quartzite and phyllite) gravel. Uncontaminated soils in previous site investigations were found to be fairly homogenous in regard to metal concentrations, i.e. expected relative standard deviation (RSD) < 50%.
4.2	Define what constitutes a sampling unit.	Sampling units will consist of: • field samples of appropriately described and logged samples which are field screened • analytical samples of the laboratory-specified sample jar quantity.
4.3	Specify temporal boundaries and other practical constraints associated with sample/data collection.	To achieve the three-month schedule for problem resolution, the field investigation should start within two weeks of the investigation plan (SAQP and commercial) being accepted. There are no site access restrictions for personnel once they are inducted and project-approved. The decision area is open with a light grass covering only and directly accessible without obstructions.
4.4	Specify the smallest unit on which decisions or estimates will be made.	The decision is to be based on the complete decision area. However, following data analysis, some form of segregation may be considered, i.e. some of the decision area may be suitable for HIL-A and some may require offsite disposal.
5	**Develop the analytic (statistical) approach** – develop a logical 'if …, then …, or …' statement that defines the conditions that would cause the decision maker to choose among alternative actions.	–
5.1	Specify the statistical parameter that characterises the population of interest, such as mean, median, maximum, 95% upper confidence limit (UCL) of the arithmetic mean or proportion.	The 95% UCL of the arithmetic mean will be the key statistical parameter. The data evaluation will include: • the 95% UCL arithmetic mean to be ≤ criterion • no individual sample to exceed 250% of the criterion • the sample standard deviation to be < 50% criterion. Additional considerations will include aesthetic requirements, including no odours or staining, no waste materials and no monolithic deposits as per NEPC 2013, B2.

No.	DQO process step	Outputs of DQOs process step
5.2	Specify the action level for the decision.	To determine if the material is suitable for the HIL-A land use, analytical action levels are to be based on the NEPC HILs (2013, B1).
		If the material is not suitable for the HIL-A land use, the material will be classified in accordance with EPA 2014c for offsite disposal.
		Samples will be held at the laboratory for additional analyses, including leachate analysis following TCLP extraction, if required.
5.3	Confirm that measurement detection will allow reliable comparisons with the action level.	Samples will be submitted to NATA-accredited laboratories. The laboratories' analytical LORs are suitably below the adopted criteria. Note: to achieve an acceptable limit of reporting for asbestos fines and fibrous asbestos, the method may not be NATA-accredited but undertaken using in-house methods for quantification.
5.4	Combine the outputs from the previous DQOs steps and develop an 'if ..., then ..., else ...' theoretical decision rule based on the chosen action level.	**If** the statistical parameters (or aesthetics) of the sampling data exceed the applicable action levels, **then** offsite disposal of the fill material will be required, **otherwise**, **if** the statistical (and aesthetic) parameters are below the applicable action levels, **then** the fill material will be determined to be suitable for a HIL-A land use.
6	**Specify performance or acceptance criteria** – specify probability limits for false rejection and false acceptance decision errors.	_
6.1	Specify the decision rule as a statistical hypothesis test.	The null hypothesis is that the material is contaminated and exceeds the adopted criteria. The alternative hypothesis is that the material is not contaminated above the adopted criteria.
6.2	Examine consequences of making incorrect decisions from the test.	Possible decision errors include:
		• the material being accepted as suitable for a HIL-A land use when it is not, thereby potentially risking human health or environmental impacts
		• unnecessary disposal of the material offsite, imposing needless financial and resource burdens on the development project and resulting in an inappropriate waste classification.
6.3	Place acceptable limits on the likelihood of making decision errors, including acceptable alpha (α) and beta (β) risk levels.	Stated hypotheses:
		• null hypothesis (H_0): the 95% UCL, and other requirements, are > the action level
		• alternate hypothesis (H_A): the 95% UCL, and other requirements, are ≤ the action level.
		Potential outcomes include Type I and Type II errors:
		• Type I error of determining the material is acceptable for the proposed HIL-A land use when it is not (wrongly rejects true H_0).

No.	DQO process step	Outputs of DQOs process step
		• Type II error of determining the material is unacceptable for the proposed HIL-A land use when it is acceptable (wrongly accepts false H_0).
		For performance criteria, the acceptable limits on the likelihood of making decision errors to be applied are:
		• alpha risk (Type I error) of $\alpha = 0.05$
		• beta risk (Type II error) of $\beta = 0.2$.
		No previously collected data are available for use, therefore acceptance criteria are not required.
7	**Optimise the design for obtaining data** – identify a resource-effective sampling and analysis design for generating data that is expected to satisfy the DQOs.	_
7.1	Document the final sampling and analysis design, along with a discussion of the key assumptions underlying this design.	To allow statistical inference, a probabilistic systematic strategy is to be adopted. As the proposed development is based on 32 residential lots of 400 m^2, 32 sample locations were selected so the density equates to one sample location per lot. Using a regular square grid size of 20 m, the grid will consist of 4 x 8 cells, with the sample locations within each cell to be selected randomly. The grid lines will be designated A to D from north to south (short axis) and 1 to 8 from west to east (long axis). The first node will be A1, through to D8. There will be 32 sample locations.
		Test pits will be excavated at each location to the underlying natural material as identified by the remnant A-horizon, at a depth of 2–3 m. Two field samples will be collected at each sample location at the surface (a depth of 0.01 m) and an approximate depth of half the total test pit depth, so there will be 64 field samples.
		Sixteen analytical samples are to be analysed initially, with the remaining field samples to be held at the laboratory. It was assumed that a relative standard deviation of about 75% could be expected and based on a maximum probable error (MPE) of between 30% and 50%, 16 samples were calculated as appropriate for analysis using the MPE method for determining the number of samples required.
		Based on the size of the decision area, this sampling design results in:
		• one sample location per forecasted residential block (400 m^2)
		• one field sample per 500 m^3
		• one analytical sample per ~2,000 m^3.
		This design is theoretically capable of detecting a minimum hotspot diameter of 23.6 m.
		The results from the first sixteen samples will be considered, and a decision on whether to utilise the remaining samples held at the laboratory will then be made. For instance, the 95% UCL should be

No.	DQO process step	Outputs of DQOs process step
		calculated and compared with the assessment criteria. If the calculation indicates that the 95% UCL is above a criterion/criteria, a calculation can be performed to determine how many samples are required to determine that the 95% UCL is below the criterion/criteria. See section 7.
7.2	Detail how the design should be implemented, together with contingency plans for unexpected events.	The field methods for sample collection, handling, and analysis (at analytical laboratories) are described in the project-level standard operating procedures (SOPs). Contingencies include collecting additional samples from material that is significantly different from the reworked natural material, and conducting additional analyses where field indicators (staining, odours, field screening results) suggest other contaminants.
7.3	Determine the quality assurance and quality control (QA/QC) procedures that are to be performed to detect and correct problems to ensure defensible results.	The required field QA, and the field and laboratory QC, are described in the project-level SOPs. These include both the data quality indicators (DQIs) and the associated measurement quality objectives (MQOs).
7.4	Document the operational details and theoretical assumptions of the selected design in the SAQP.	Theoretical assumptions include: • surficial impacts from overland flow from the adjacent facility and burial of wastes are the modes of contamination expected • the material is relatively homogenous • the remnant A-horizon will be readily discernible from buried grass and organic soil.

The resulting detected metal data from the 'implementation' of this investigation is summarised in **Table 8**.

Table 8   Summary of analytical results – metals in soil (mg/kg)

Sample/descriptor	Arsenic	Chromium	Copper	Lead	Nickel	Zinc
LORs	5	2	5	5	2	5
Analytical						
Analytical sample B2-01	103	12	34	20	18	11
Analytical sample B2-02	50	21	30	7	2	10
Analytical sample D2-01	43	26	83	17	14	35
Analytical sample D2-02	9	10	29	14	5	12
Analytical sample A4-01	203	4	260	18	12	232
Analytical sample A4-02	54	5	55	17	9	41
Analytical sample C4-01	341	19	401	133	7	543
Analytical sample C4-02	34	17	46	16	10	13
Analytical sample B6-01	71	18	24	14	5	9
Analytical sample B6-02	14	6	8	17	12	5
Analytical sample D6-01	62	11	51	15	3	36
Analytical sample D6-02	6	4	18	16	24	10
Analytical sample A8-01	27	17	61	16	4	24
Analytical sample A8-02	7	10	38	20	13	10
Analytical sample C8-01	24	15	39	12	6	8
Analytical sample C8-02	13	16	17	14	19	7
Descriptive statistics						
Number of samples	16	16	16	16	16	16
Number of detects	16	16	16	16	16	16
Percentage non detects	0%	0%	0%	0%	0%	0%
Maximum	341	26	401	133	24	543
Third quartile	64.3	17.3	56.5	17.3	13.3	35.3
Median value	38.5	13.5	38.5	16.0	9.4	11.5
First quartile	13.8	9.0	27.8	14.0	5.2	9.8
Minimum	6	4	8	7	2	5
Arithmetic average	66.3	13.2	74.6	22.9	10.2	62.9
Geometric average	35.2	11.4	43.5	17.3	8.3	20.0
Mode	–	10	–	17	12	10
Variance	7,792.2	42.4	10,988.8	872.1	39.7	19,410.1
Standard deviation	88.3	6.5	104.8	29.5	6.3	139.3
Coefficient of variation (CV)	1.3	0.5	1.4	1.3	0.6	2.2
Inferential statistics						
Standard error of the mean (SEx̄)	22.1	1.6	26.2	7.4	1.6	34.8

Sample/descriptor	Arsenic	Chromium	Copper	Lead	Nickel	Zinc
Relative standard deviation (RSD)	133.1%	49.4%	140.5%	129.1%	61.9%	221.6%
Margin of error (MoE)	47.0	3.5	55.9	15.7	3.4	74.2
Maximum probability error (MPE)	70.9%	26.3%	74.9%	68.8%	33.0%	118.1%
95% UCLx̄ two-sided Student's t	113.4	16.7	130.5	38.6	13.5	137.1
95% UCLx̄ one-sided Student's t	105.0	16.0	120.5	35.8	13.0	123.9
ProUCL determination	120.5	16.0	135.2	55.1	13.0	214.7
Method recommended	Gamma	Student's t	H-UCL	Chebyshev	Student's t	Chebyshev
Criteria and number of samples						
HIL-A land use (NEPC 2013, B1)	100	100	6,000	300	400	7,400
Number of samples CRV method	43.9	1.4	1.4	1.4	1.4	1.4
Number of samples CRV method	44	2	2	2	2	2
Number of samples MPE method	15	18	16	16	14	15

Notes

LORs = limits of reporting

For determination of descriptive statistics, see Appendices A to D of Part 2 of these guidelines (*Sampling design part 2 – interpretation*).

SEx̄ – see Appendix I of Part 2 of these guidelines.

RSD – see Appendix A of Part 2 of these guidelines.

MoE – see Appendix I of Part 2 of these guidelines.

MPE – see Appendix I of Part 2 of these guidelines.

For determination of lower confidence limit (LCL) and upper confidence limit (UCL), see Section 5 of Part 2 of these guidelines.

ProUCL = USEPA's ProUCL, Version 5.1

For determination of number of samples, see Appendices E and F of this document.

Appendix C: Determining sampling grids for hotspot detection

This appendix provides the methods for determining the required grid size, for square grids, to detect hotspots of a specified size. The method for determining the approximate number of sampling locations, based on the hotspot shape and size, is also provided. However, as the number of sampling locations required is in part based on the geometry of the site or decision area, the actual number of sampling locations required is dependent on applying the specified grid size to the actual site or decision area.

Determination

For determining grid size:

Equation 1

$$G = \frac{r}{k}$$

For determining the number of sampling locations:

Equation 2

$$n = \frac{A}{G^2}$$

For determining the critical size of hotspots:

Equation 3

$$r = k \times G$$

Where:

- **G** is the grid size, that is, the distance between nodes of grid
- **r** is the radius of a circular hotspot (for intermediate-shaped and elliptical hotspots, halve the length of the major axis)
- **k** is a statistical constant, dependent on the shape of the hotspot and the required confidence level
- **n** is the number of sampling locations
- **A** is the area of the site or decision area.

The values for k at 95% confidence level were determined from Figure 10.3 in Gilbert 1987, as:

- 0.59 for circular hotspots (ratio is 1:1)
- 0.69 for intermediate shaped hotspots (ratio is 4:3)
- 0.9 for elliptical hotspots (ratio is 2:1).

Gilbert 1987 notes that for elliptical targets, the curves in Figure 10.3 are 'average curves over all possible orientations of the target relative to the grid'.

To determine the required grid size, choose the expected size (r) and shape (k) of the hotspot, then determine the required grid size from Equation 1. Use Equation 2 to determine n, the number of sampling locations required, and Equation 3 to determine r, the minimum hotspot size that can be detected.

If the contaminant is known or suspected to exhibit periodic spatial variations, the sampling pattern should be oriented so it will not be in or out of phase with the known or suspected periodic spatial variations.

Appendix D: Summary of existing guidance for sample design

Table 9 summarises sampling design information from guidance made and approved under Section 105 of the *Contaminated Land Management Act 1997*, and from other guidance documents. The specified guidance should be referred to for details of the sampling strategies, locations, sampling densities and contaminants of potential concern (CoPC).

Table 9   Existing guidance for sampling design

Situation or land use	Guidance	Medium	Sampling design information
All land uses	National Environment Protection Council (NEPC) 2013, *National environment protection (assessment of site contamination) amendment measure 2013 (No. 1)*, Schedule B2, National Environment Protection Council, Canberra.	Soil, soil gas and groundwater	Provides judgmental and probabilistic sampling design information for various media, including stockpiles.
Banana lands	Environment Protection Authority (EPA) 1997, *Contaminated sites: guidelines for assessing banana plantation sites*, EPA 97/37, NSW EPA, Sydney.	Soil	Provides information to investigate and assess potential contamination on current and former banana growing lands, including CoPC. A systematic/grid-based sampling strategy is recommended, with variable sampling densities based on the former use, the current land use, and the stage of the investigation. A judgmental sampling design is recommended during validation of any excavations.
Cattle-tick dip sites	McDougall KW & Macoun TW 1996, *Guidelines for the assessment and clean up of cattle tick dip sites for residential purposes*, NSW Agricultural in conjunction with CMPS&F Environmental, Wollongbar NSW.	Soil	Provides information to assess and remediate sites containing former cattle tick dips, including an overview of the CoPC and the areas of highest potential contamination. A stratified systematic sampling design is recommended, with sampling of the sub-area based on the likelihood for contamination from past use.

Situation or land use	Guidance	Medium	Sampling design information
Excavated natural material (ENM) order	Environment Protection Authority (EPA) 2014a, Resource Recovery Order under Part 9, Clause 93 of the Protection of the Environment Operations (Waste) Regulation 2014, *The excavated natural material order 2014*, NSW EPA, Sydney.	Soil	Provides information to allow the adequate assessment and classification of ENM for resource recovery. The order stipulates sampling strategies based on in-situ or stockpiled material. For in-situ material, a systematic/grid-based sampling strategy at specified depth intervals is required. When stockpiled, the number of samples by volume is specified, noting judgement must be used to ensure that samples taken are representative of the material.
Ground gas	Environment Protection Authority (EPA) 2020a, *Assessment and management of hazardous ground gases: contaminated land guidelines*, EPA 2019P2047, NSW EPA, Sydney.	Soil gas	Provides information to assist with the investigation of sites with potential hazardous bulk and trace ground gases. Includes judgmental sampling design information to apply for site specific scenarios.
Groundwater	Department of Environment and Conservation (DEC) 2007, *Contaminated sites: Guidelines for the assessment and management of groundwater contamination*, DEC 2007/144, DEC NSW, Sydney.	Groundwater	Provides information to conduct groundwater investigations, including a description of relevant concepts to allow for an adequate sampling design program to be developed.
Gasworks	Department of Environment and Conservation (DEC) 2005b, *Information for the assessment of former gasworks sites*, DEC 2005/237, DEC NSW, Sydney.	Soil and groundwater	Provides information relating to former gasworks sites and the potential for contamination of site areas and CoPC. Recommends stratifying the site and using a systematic sampling design to ensure sufficient sampling density for each area of concern. Describes that groundwater monitoring well locations should consider the site-specific complexities of the hydrogeology.

Situation or land use	Guidance	Medium	Sampling design information
Land farming	Environment Protection Authority (EPA) 2014b, *Best practice note: lLandfarming*, EPA 2014/0323, NSW EPA, Sydney.	Soil	Provides information on best practice land farming techniques and recommends a systematic sampling design that is adequate to provide a statistically reliable result.
Orchards and market gardens	Department of Environment and Conservation (DEC) 2005a, *Contaminated sites: guidelines for assessing former orchards and market gardens*, DEC 2005/195, DEC NSW, Sydney.	Soil	Provides information relating to former orchards and market gardens sites, the potential for contamination of specific site areas and relevant CoPC. Recommends a systematic grid-based sampling plan across the cultivated areas of the site, targeting the surface soils, with a higher sampling density for areas where localised contamination is likely to have occurred i.e. chemical storage sheds and tractor turning circles.
Resource Recovery Order/exemptions	Environment Protection Authority (EPA) 2018, *Guidelines on resource recovery orders and exemptions: For the land application of waste materials as fill*, EPA 2017/P0392, NSW EPA, Sydney.	Soil and fill	Provides information for Resource Recovery Order/exemptions to allow beneficial reuse of waste products such as fuel, fill and fertiliser. The guideline provides a minimum number of samples which must be collected and specifies that the 'sampling plan must have a clear, defensible rationale', implying the use of probabilistic systematic sampling designs.
Service stations and underground petroleum storage systems (UPSS)	EPA 2020c, *Guidelines for implementing the Protection of the Environment Operations (Underground Petroleum Storage Systems) Regulation 2019*, EPA 2020/P2700, NSW EPA, Parramatta. **and** UPSS fact sheets on the <u>EPA website</u>.	Soil and groundwater	Provides information to investigate and assess contamination at service stations or locations with UPSS. Judgmental sampling design is recommended, targeting soil and groundwater in areas of infrastructure and known contamination.

Situation or land use	Guidance	Medium	Sampling design information
Stockpiles	National Environment Protection Council (NEPC) 2013, *National environment protection (assessment of site contamination) amendment measure 2013 (No. 1)*, Schedule B2, National Environment Protection Council, Canberra.	Soil and fill	Section 7.5 of the NEPC (2013, B2) provides information for assessing stockpiles of homogenous soil or fill of ≤ 200 m^3. It recommends a minimum number of samples to undertake an initial assessment of a stockpile, with either a judgmental or probabilistic sample design recommended based on the specific circumstance.
Surface water	Australian and New Zealand Environment and Conservation Council (ANZECC) and Agriculture and Resource Management Council of Australia and New Zealand (ARMCANZ) 2000, *Australian and New Zealand guidelines for fresh and marine water quality*, paper no. 4, ANZECC and ARMCANZ, Canberra. Available at: www.waterquality.gov.au/anz-guidelines.	Surface waters	Provides detailed guidance for the management and assessment of waters in Australia and New Zealand. Information is provided on how to develop an appropriate surface water sampling program.
Vapour intrusion	Environment Protection Authority (EPA) 2020a, *Assessment and management of hazardous ground gases: contaminated land guidelines*, EPA 2019P2047, NSW EPA, Sydney **and** ,	Soil gas and volatiles	Provides information for assessment of sites that have potential vapour intrusion issues. Judgmental and probabilistic sampling design information for the various vapour intrusion investigation methods, including conceptual information to determine the number of sample locations and frequency.
Vertical mixing of soil on former broadacre agricultural land	Environment Protection Authority (EPA) 1995, *Contaminated sites: guidelines for the vertical mixing of soil on former broad-acre agricultural land*, EPA 2003/28	Soil	Provides information regarding use of vertical mixing techniques of former agricultural land. No specific sampling design information is provided beyond sample depths.

Appendix E: Determining the number of samples by the CRV method

The number of samples needed to show that the mean concentration of a contaminant is below a defined action level or criteria can be determined using the combined risk value (CRV) method. This method can be used for sites or decision areas for all media, where probabilistic sampling has been undertaken.

The determination derived from the Student's t-test formula for hypothesis testing, with the alpha (α) value for a Type I error, or false rejection of the hypothesis, and the beta (β) value for a Type II error, or false acceptance of the hypothesis, is used to determine the CRV. In assessing site contamination, the null hypothesis (H_0) is always that the contaminant concentrations exceed the action levels or criteria, and where H_0 is not rejected, there is only a potential for a Type II or false acceptance error rate, and this sample size formula can be used to determine if the error rate has been satisfied.

This method can be used to design a sampling program, either using previous data or estimates to determine x̄ and s, or retrospectively to demonstrate sufficient statistical power or otherwise. Where the determination results in low values of n, including ≤ 1, this suggests that the minimum detectable difference Δ (uppercase Greek letter delta) is overly large, and additional statistical analysis is required to determine or justify the number of samples.

Determination

1. The number of samples using the CRV method is determined by:

$$n = \frac{\left(Z_{1-\alpha} + Z_{1-\beta}\right)^2 * s^2}{(C_s - \bar{x})^2}$$

Where

n   number of samples

Z   standard normal distribution (z curve)

Z_{1-a}   Z value for α

$Z_{1-\beta}$ Z value for β

C_s   criterion/action level

x̄   sample mean

s   sample standard deviation.

2. The risk values are selected for $Z_{1-\beta}$ and $Z_{1-\alpha}$ from USEPA 1989.

When comparing to action levels or criteria, the recommended values are 0.05 α risk and 0.2 β risk, corresponding to confidence levels of 95% and 80% respectively. Using a 0.05 α risk value of 1.645 and a 0.2 β risk value of 0.842, the CRV is 6.2.

Where increased certainty is required, such as determining if costly remedial works are necessary, consultants are encouraged to examine the use of more conservative β values, which will result in an increased CRV. As β corresponds to the risk of falsely accepting the H_0, that is, that the site is contaminated, further sampling reduces the chance of Type II errors, which could potentially lead

to the rejection of H$_0$, making the decision that the site is not contaminated. Generally, the cost of unnecessary remediation will far outweigh the cost of the additional sampling and analysis.

Worked example

The metals data in mg/kg from Table 8 in Appendix B: Data-quality objectives: worked example is used in this example to confirm that the number of samples collected for arsenic (As) and chromium (Cr) is appropriate, i.e. that the statistical power of the test is sufficient.

1. Select the confidence level and the power of the test. For $\alpha = 0.05$ and $\beta = 0.2$, the solution is:

$$n = \frac{6.2 * s^2}{(C_s - \bar{x})^2}$$

2. The number of samples required using the CRV method is determined at $\alpha = 0.05$ and $\beta = 0.2$ as follows.

Arsenic

3. For As, $\bar{x} = 66.3$, $s = 88.3$ and HIL-A = 100 mg/kg:

$$n = \frac{6.2 * 88.3^2}{(100 - 66.3)^2}$$

$$n = 42.5$$

Rounding up to the nearest whole number, 43 samples are required to characterise the site or decision area for As, based on the large standard deviation. As the maximum concentration of As exceeds HIL-A by more than 250%, additional investigation is required to further characterise the distribution of As. The large number of samples required, and the large value of s, suggests that further characterisation should seek to segregate the decision area into different sub-populations, either in plan or by depth, for the design of further investigations and consideration of remedial options.

Chromium

3. For Cr, $\bar{x} = 13.2$, $s = 6.5$ and HIL-A = 100 mg/kg (Cr^{6+}):

$$n = \frac{6.2 * s^2}{(C_s - \bar{x})^2}$$

$$n = \frac{6.2 * 6.5^2}{(100 - 13.2)^2}$$

$$n = 0$$

No samples are required to characterise the site or decision area. This is not surprising, based on the small standard deviation and mean. However, with no samples it is not possible to estimate the true values of the critical parameters of the contaminant distribution, such as $\bar{x}$ and s. Rather, the use of the CRV method and the derivation of n less than the number of samples collected, suggests that the false rejection (α) error rate has been satisfied, and that in the case of Cr, it is reasonable to reject H_0 (i.e. that the site or decision area is contaminated with Cr).

Appendix F: Determining the number of samples by the MPE method

The number of samples needed to show that the average concentration of a contaminant is within a specific range, such as a confidence interval, can be determined using the maximum probable error (MPE) method. This can be thought of as a specified statistical precision around a point estimate and can be used for any medium and any probabilistic sampling design.

This method can be used when addressing estimation problems as defined within the data quality objectives (DQOs) process, as it allows a desired precision to be specified outside a strict hypothesis-testing framework.

This method uses the margin of error (MoE), the standard deviation (s), and a critical value at a specified confidence level. The MoE can be thought of as the 'radius' to, or half the width of, the diameter of the confidence interval. Initially, $Z_{1-\alpha/2}$ is used for a first determination, the result of which defines the degrees of freedom for selection of a value for $t_{1-\alpha/2,n-1}$. Subsequent iterations are conducted until the number of samples calculated stabilises.

In the case of the MPE method, as the equation reduces to n = n, it cannot be used to retrospectively demonstrate sufficient sampling specifically, but provides a guide to an appropriate number of samples based on the variability of the data (standard deviation), and the required precision of the data MoE. If the data shows too large an MPE (> 35–50%) for reasonable relative standard deviations (RSDs) (65–150%), the data is probably not sufficiently precise.

Table 10 shows various values for n, calculated using USEPA 2015: they are illustrated in Figure 11. As the RSD increases, and higher precision (lower MPE) is required, the number of samples required increases.

Determination

1. The number of samples is calculated by the MPE method as:

$$n = Z_{1-\alpha/2}{}^2 * \frac{s^2}{MoE^2}$$

Where:

n   number of samples

$Z_{1-\alpha/2}$ Z from the standard normal distribution

$t_{1-\alpha/2,n-1}$ critical value

s   sample standard deviation

MoE   margin of error (= $t_{1-\alpha}$ * $SE_{\bar{x}}$)

$SE_{\bar{x}}$ standard error of the mean (= s/√n).

2. The MoE and s can be standardised as relative values by dividing by x̄, giving the maximum probable error (= MoE/x̄) and the relative standard deviation (RSD) (= s/x̄), which is also known as the coefficient of variation (CV). Using the standardised MPE method, the required number of samples is calculated by:

$$n = t_{95\%}^2 * \frac{RSD^2}{MPE^2}$$

Either method can be used as long as the variables are consistent, that is, s and MoE are expressed in mg/kg, or RSD and MPE are given as percentages.

Worked example

The metals data in mg/kg from Table 8 is used in this example to determine if sufficient samples have been collected for copper (Cu), both at the surface and at depth.

1. The number of samples required is initially determined using Z, as:

$$n = Z_{1-\alpha/2}^2 * \frac{s^2}{MoE^2}$$

2. For the surface samples, at a 95% confidence level, Z = 1.96, s = 137 and MoE = 114.5.

$$n_1 = 1.96^2 * \frac{137^2}{114.5^2}$$

$$n_1 = 5.5$$

3. Rounding to the next whole number, 6, the degrees of freedom is 5. Using $t_{1-\alpha/2,n-1}$ = 2.571, the next determination is:

$$n_2 = t_{1-\alpha/2}^2 * \frac{s^2}{MoE^2}$$

$$n_2 = 2.571^2 * \frac{137^2}{114.5^2}$$

$$n_2 = 9.5$$

4. This process is continued until at n_4, n stabilises at 8, which was the number of samples collected.

5. The same process is used for the samples collected at depth, which also stabilises at n = 8, which was the number of samples collected.

As discussed, this approach cannot be used to confirm retrospectively if an appropriate number of samples was collected. However, by examining the RSD and the MoE achieved by the number of samples that were collected, it can be determined if enough samples were collected to meet the project requirements for the desired quantity and quality of the data.

In the present example, we can compare the results for the material collected from the surface with those for the material collected from depth.

The RSD is 115% for the surface sample results and 52.8% for the deeper sample results, and the MPE is 96.1% for the surface material and 44.2% for the deeper material. Table 10 shows that these values give a sample number of 8 (by interpolation).

For the surface material with an RSD of 115%, to achieve an MPE of 50%, 23 samples would be required (by interpolation).

In the case of the deeper material, based on the homogenous nature of the material, as indicated by the low RSD (~50%) and the precision of the data (MPE of ~45%), it is likely that the dataset would be suitable for a decision.

For the surface results, Cu is well below the HIL-A of 6,000 mg/kg, so it would probably be considered that sufficient samples have been collected from the surface fill to make a decision. However, where the dataset exhibits high RSDs and approaches the criteria or action levels, the MPE method provides a tool for assessing the quantity and quality of the data for making decisions. Based on the Cu results, the material is suitable.

.

Table 10   **Number of samples (n) required to estimate mean, based on the MPE method**

RSD %	Maximum probable error %											
–	10	15	20	25	30	35	40	45	50	55	75	100
10	6	4	3	3	3	3	3	3	3	3	2	2
15	11	6	5	4	3	3	3	3	3	3	3	3
20	18	9	6	5	4	4	3	3	3	3	3	3
25	26	13	8	6	5	4	4	4	3	3	3	3
30	37	18	11	8	6	5	4	4	4	4	3	3
35	49	23	14	10	8	6	5	5	4	4	3	3
40	64	30	18	12	9	7	6	5	5	4	4	3
45	80	37	22	15	11	9	7	6	6	5	4	3
50	98	45	26	18	13	10	8	7	6	6	4	3
55	119	54	31	21	15	12	10	8	7	6	4	4
60	141	64	37	25	18	14	11	9	8	7	5	4
70	191	86	49	33	23	18	14	12	10	9	6	4
80	248	112	64	42	30	22	18	15	12	11	7	5
90	314	141	80	52	37	28	22	18	15	13	8	6
100	387	173	98	64	45	34	26	21	18	15	9	6
110	467	209	119	77	54	40	31	25	21	18	11	7
120	556	248	141	91	64	48	37	30	25	21	12	8
130	652	291	165	106	75	55	43	34	28	24	14	9
140	755	337	191	123	86	64	49	40	33	27	16	10
150	867	387	219	141	98	73	56	45	37	31	18	11
175	1,179	525	297	191	133	98	76	61	49	41	23	14

| **200** | 1,539 | 685 | 387 | 248 | 173 | 128 | 98 | 78 | 64 | 53 | 30 | 18 |

RSD = relative standard deviation (s/x̄, where s is the sample standard deviation and x̄ is the sample arithmetic mean).

MPE = maximum probable error (MoE/x̄, where MoE is the margin of error (= $t_{95\%}$ * s/√n)).

Shaded values represent the general range of n required for characterising homogenous material within the same decision area.

Figure 11   Number of samples (n) required to estimate mean, based on the MPE method
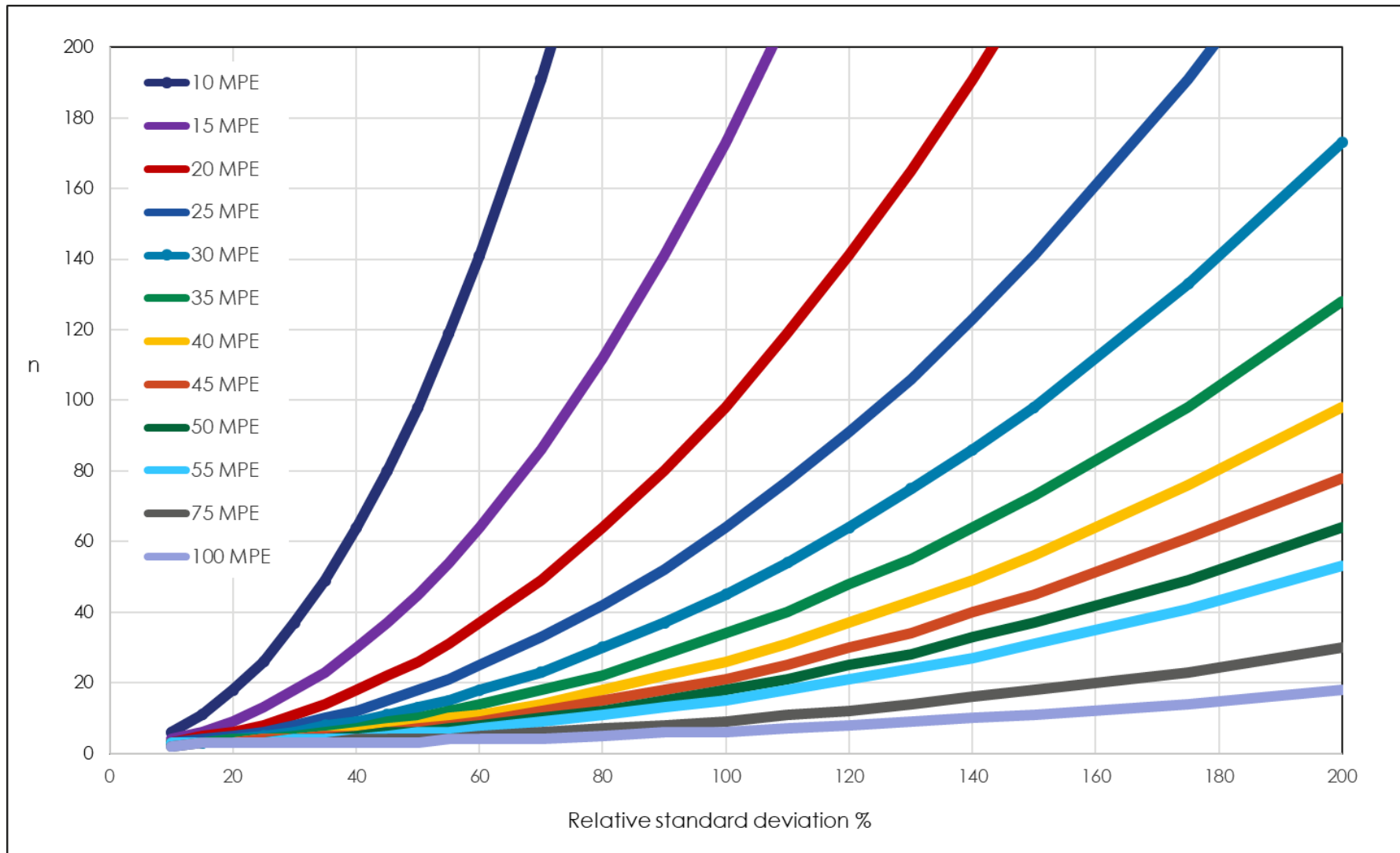
Source: Easterly Point Environmental Pty Ltd

Appendix G: Further methods for consideration

The following section provides a summary of some investigation methods commonly used by USEPA, and also introduces geospatial statistics.

Triad approach

The Triad approach is a decision-making framework USEPA developed to manage decision uncertainties in environmental data. It draws on new advances in science and technology and allows projects to proceed rapidly and data to be collected cost-effectively (Crumbling et al. 2004).

Traditionally, site characterisation has proceeded through many stages of investigation, emphasising the assessment of analytical data based on the relationship between data quality and analytical quality. However, this approach can be quite static, repetitious, time-consuming and expensive. In contrast, the Triad approach focuses on a more adaptive approach to site characterisation, using real-time decision-making tools to guide field activities, creating more flexibility and reducing overall costs and resources.

Preliminary evaluations suggest that incorporating the Triad approach into the decision-making framework can save up to 50% of the cost of more traditional approaches to site characterisation (Crumbling 2001).

The Triad approach has three main elements. While these elements are not new concepts in the site-investigation process, what **is** new is how they are synthesised to 'plan, implement and improve data collection from contaminated sites' (Clements et al. 2009). The three elements are:

Systematic planning

This is the most important element of the Triad approach and ensures high decision confidence. Systematic planning includes applying the data quality objectives (DQOs) process and developing a conceptual site model (CSM). These planning tools can then be used to inform stakeholders by providing a clear understanding of the site, the uncertainties identified, and the required data objectives. This stage of the process also allows for stakeholder involvement and consensus regarding the desired project outcomes, including any end goals and exit strategies, which are clearly defined before field work commences. This contrasts with the traditional approach, where a decision is made after the site investigation has been conducted, based solely on the results of analytical data (Crumbling et al. 2004).

Dynamic work strategies

These strategies enable projects to be completed much faster and at considerably less cost than using more traditional, static work strategies. Work planning documents are prepared to allow for flexibility in project planning as data from field measurements becomes available. For example, a sampling and analysis quality plan (SAQP) may include contingencies that allow field work to be modified, even while it is still occurring (Clements et al. 2009), enabling the CSM to be a 'dynamic' document that can be refined as more site information and data become available.

Real-time measurement systems

By reviewing field screening data and analytical data in real-time, decisions such as remediation strategies and adaptive sampling plans, for example, revised sampling locations, sample quantities or analytical strategy, can be made while the fieldwork team is still on-site. This element of the Triad approach also allows for the data to be shared among all stakeholders as soon as it is

generated, creating transparency which helps to establish trust and good working relationships with regulators and stakeholders, while also informing the decision-making process (Crumbling et al. 2004).

Traditionally site investigations have focused on the quality of the analytical data, that is, the analytical data are considered to be 'definitive data' and of a 'high quality', while real-time data and field screening methods are considered to generate 'screening data', that is, 'inferior' quality data (Crumbling 2001). In fact, the quality assurance conducted as part of the Triad approach can be 'more relevant and supportive of defensible project decisions than [that done] under traditional scenarios' (Crumbling et al. 2004). The Triad process also improves project quality by recognising the potential impacts of uncertainties in site heterogeneity, which are often overlooked in traditional site assessments and project planning. Figure 12 shows examples of real-time measurement technologies.

Figure 12   Real-time measurement technologies
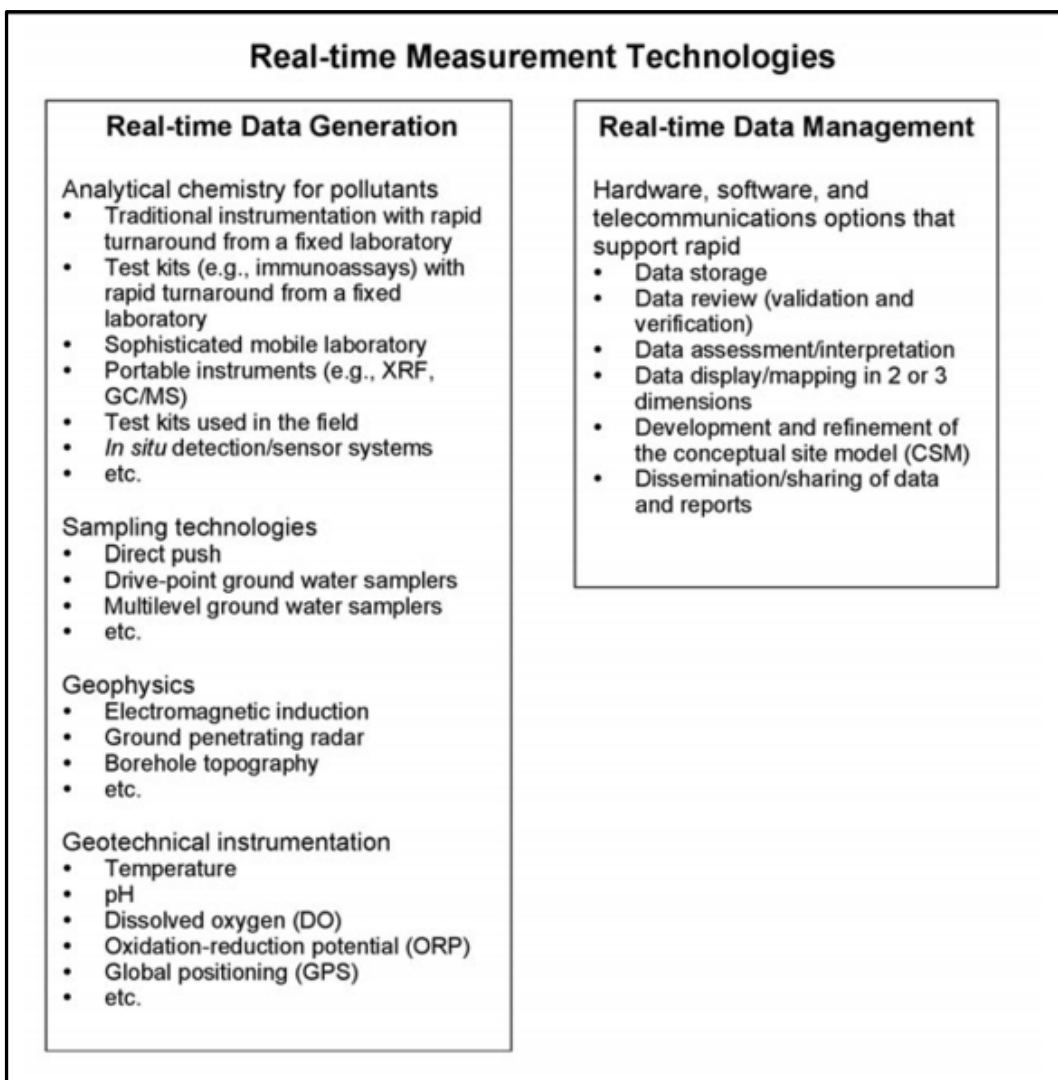
Source: Crumbling et al. 2004

Summary

Staged investigation processes came into use at a time when technology, science and consultant experience in relation to contaminated sites were limited. Now due to new technologies and science, consultants better understand contamination scenarios, environmental fate and transport

processes. The Triad approach increases decision confidence by providing an adaptive approach that focuses on real-time decision making to guide field activities and the development of a dynamic CSM, and can be a more cost- and resource-effective alternative to the traditional multi-stage investigation process. However, for the Triad approach to be effective, all its key concepts must be used.

Geospatial statistics

Geostatistical data analysis is based on **bivariate statistics** theory. Bivariate statistics allow for the assessment of the relationships between two variables. The bivariate statistical approach for assessment of concentration data was developed in the 1960s for use in the minerals exploration industry (Krige 1981). This approach examines concentrations and variations in relation to their spatial distribution. Given the similar objectives for the assessment of contamination at a site, this method has application in the assessment of contaminant concentration and location data. Bivariate geostatistics assess not only the distribution of the contaminant but also the spatial variance in concentration (Goovaerts 1997; Webster & Oliver 2001; Nielsen & Wendroth 2003).

Variograms

The **variogram** is the basic tool of geostatistics (Royle 1980) and expresses the spatial correlation between adjoining samples. Variograms are essentially scatter plots of distance between sample locations and variance of sample values to establish whether there is a predictable change in variance with distance.

A variogram is constructed by calculating the mean squared difference (variance) between sample values over incremental sample spacing. Methodologies for calculating the mean squared difference and construction of variograms are outlined in the references (Henley 1981, Krige 1981, Rendu 1981 and Royle 1980) so are not outlined in detail here.

Variograms present key geostatistical properties, which are shown in Figure 13. The range of influence is the distance over which the samples values are related. The total variance of the samples can be split into a random and a spatial component. Random variance is also known as the **nugget effect**.

Development of variograms generally requires specialist software or spatial data assessment software.

Figure 13   Features of a variogram
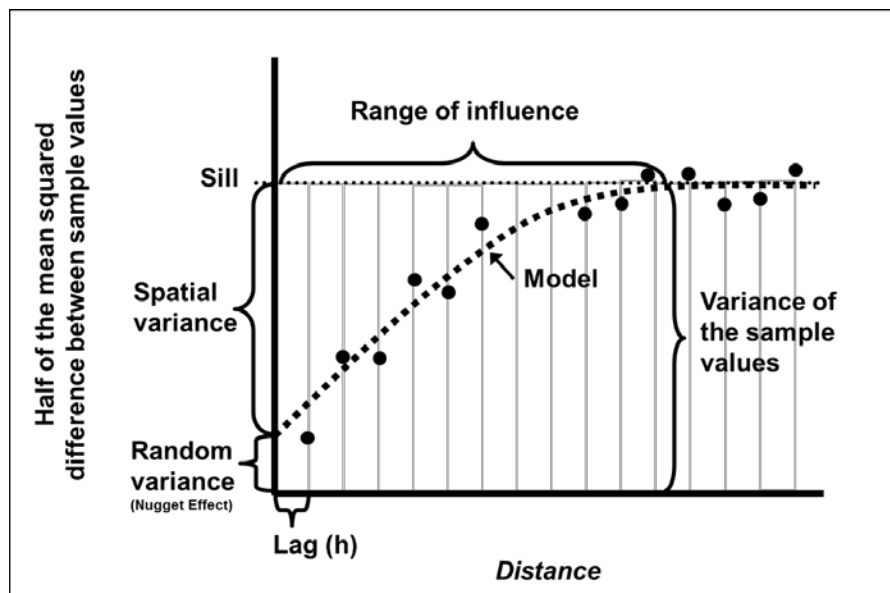
Source: Beck, Mikov and Curtis 2004

Spatial interpolation methods

The data from the variogram is interpreted using the **kriging** method (Krige 1981; Rendu 1981), which is a weighted linear estimation technique (Royle 1980). Kriging provides an estimate of a value at a given location where no site-specific measurement has been made (Henley 1981). It can be used to predict the value (such as a concentration) at a location for which no data exists, or to predict the confidence in the interpretation by using **indicator kriging** (Krige 1981; Rendu 1981; Isaaks & Srivastava 1989).

Kriging is the most reliable interpolation method for predicting values away from locations that were sampled, due to the absence of bias commonly associated with other interpolation methods.

Benefits of spatial geostatistics

The advantages of spatial geostatistics over the univariate approaches commonly used include:

- not relying on sample collection being unbiased, therefore allowing for a single statistical method for all sampling data
- assessing the spatial and random contribution to concentration variation
- providing a method to establish when a site is adequately characterised
- providing the most reliable interpolation method for spatial concentration data
- providing a reliable method for probabilistic mapping of occurrence of contamination
- providing a reliable method for probabilistic evaluation of volumes of contaminated media.

Spatial geostatistics can be considered for sites where more than 10 sampling locations have been completed, as the method requires a reasonable number of samples to be applied effectively.

Example of application

The dataset used in this example was generated by a staged investigation at a large parkland and sporting ovals. The site was suspected to have been filled with soil from a gasworks. Data from the first stage of the investigation was used to develop a variogram for the polycyclic aromatic hydrocarbon (PAH) data to assess whether there were signs of a spatial relationship. Figure 14 shows the variogram derived from the first 26 samples analysed.

Figure 14 Variogram for PAHs from the first stage of the investigation
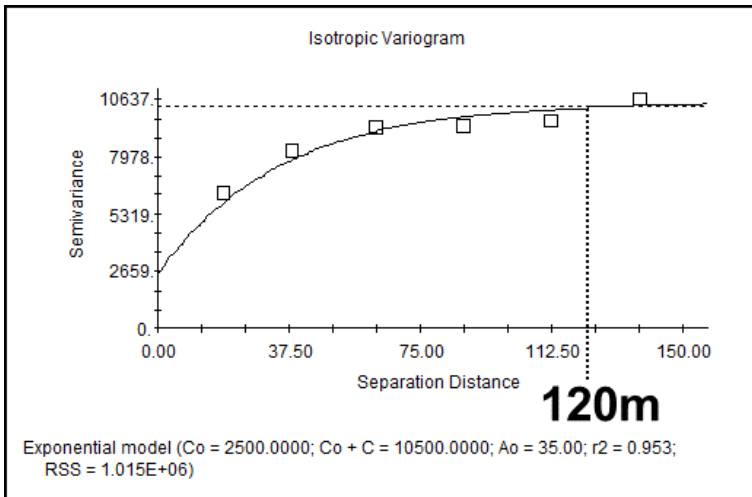
Source: Beck, Mikov and Curtis (2004)

Figure 14 shows that around 70% of the data variance is spatial and shows a range of around 120 m, while around 30% is random. Figure 14 informed the design of a second sampling round by using indicator kriging to identify areas of low confidence in the spatial distribution of contamination. A second sampling round identified 10 sampling locations that would help improve the confidence in spatial interpretation. The data generated by the initial and second sampling round was used to develop a second variogram, which is shown below.

The second variogram (Figure 15) showed a notable decrease in the random variance to around 10% of the total variance, while the range remained relatively similar.

Figure 15 Second variogram for PAHs
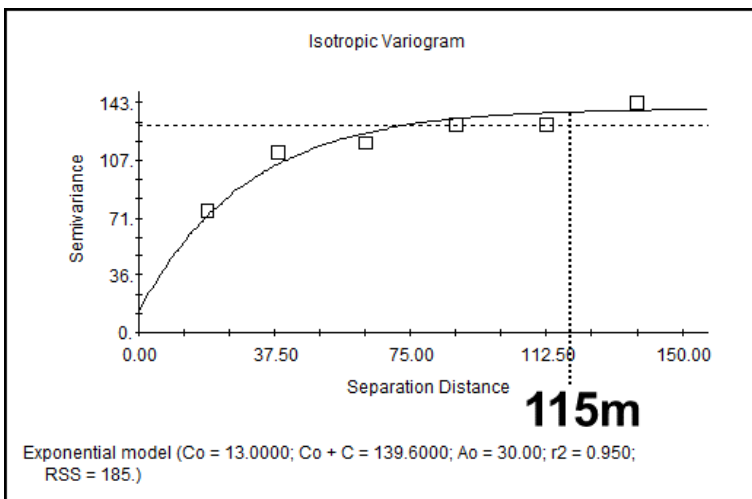
Source: Beck, Mikov and Curtis 2004

The second variogram was used to design the third sampling round, in which samples were collected from a further 12 locations. The variogram developed from the dataset after the third sampling round is shown in Figure 16.

Figure 16   Third variogram for PAHs

Isotropic Variogram

Exponential model (Co = 13.0000; Co + C = 139.6000; Ao = 30.00; r2 = 0.950; RSS = 185.)
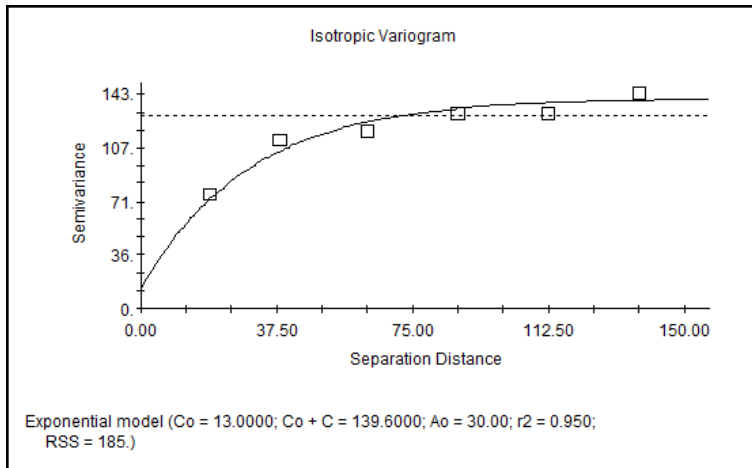
Source: Beck, Mikov and Curtis 2004

The third variogram was almost identical to the second, suggesting that further sampling would not improve characterisation of the site. However, the indicator kriging showed **that over 85% of the site was covered, at a confidence level of 80% or higher.**

Appendix H: Guide for non-technical assessors of sampling design in contaminated land reports

The purpose of this Appendix is to enable non-technical readers to understand whether a given sampling strategy is appropriate, and to provide some considerations when interpreting contaminated sites reports.

Contaminated land reports must comply with EPA guidelines

Contaminated land reports published after April 2020 must comply with the Consultants reporting on contaminated land: contaminated land guidelines, 2020, EPA 2020P2233, NSW EPA, Parramatta (EPA, 2020b). These are statutory guidelines made by the EPA under section 105 of the CLM Act. Contaminated land reports published before April 2020 were required to comply with the earlier version of the above guidelines.

The contaminated land report should state that it has been prepared in accordance with the reporting guidelines above.

The reporting guidelines include checklists that detail the information required for each section of a contaminated land report. These may assist the assessor when reviewing a report.

Determining sampling design appropriateness

To assist in determining the appropriateness of the chosen sampling design, the report must address the following:

1. The area and boundaries of the site must be known and stated.
2. A site history must have been prepared, and the report must identify if fill of unknown origin is known or suspected to be at the site.
3. A site inspection must have been carried out.
4. The report must include a conceptual site model (CSM).
5. The assessment criteria must be stated. These are also called 'investigation levels' or 'screening levels' and the analytical results must be compared against these values.
6. Consultants may calculate the 95% upper confidence limit (95% UCL) for a set of results to compare against the assessment criteria. For a definition of 95% UCL, see the Glossary.

Items 1 – 4 must be prepared in accordance with the checklists available in EPA (2020b).

Sampling regime being used

Reports that contain results of sampling and analysis must clearly state what sampling regime is being used, and this must be supported by evidence based on the site history. Three different sampling regimes are described in Section 5.2 of this document:

- **Judgmental** – targeting areas of known or suspected contamination such as underground storage tanks (USTs) or areas beneath previous structures
- **Systematic** – grid across the whole of the site or a sub-area
- **Stratified** – a mix of judgmental and systematic regimes applied to different site areas, depending on the site history.

If a **judgmental regime** is used, the report should justify why specific locations have been chosen and identify what the sampling is targeting.

If a **systematic regime** is used, the report should contain details of which sampling approach has been used and why, for example, whether the sampling locations have been chosen based on a square, offset square, random, triangular or herringbone grid pattern.

If a **stratified regime** is used, the report must clearly explain which areas have been subject to what sort of sampling, including appropriate justification for all decisions.

For an overview of the process, refer to Flow Chart 1: Choosing a sampling regime and Flow Chart 2: Choosing a sampling regime where there have been potentially contaminating activities.

Flow Chart 1: Choosing a sampling regime

Consultants complete a site history review and a site inspection in accordance with *Consultants reporting on contaminated land (EPA 2020)*.

······▶ If the site history is incomplete.

Do the complete site history and site inspection indicate any potential for contamination, including known or suspected fill of unknown origin[1]?

NO

YES

Is the fill of unknown origin the **only** potential source of contamination?

NO

YES

Choose a sampling regime from Section 5.
Prepare SAQP for confirmatory sampling.
Record observations during site walkover to support conclusions.

Choose a sampling regime from Section 5[2]. Areas subject to grid sampling will need to meet or exceed Table 2, Section 5.2.5 sampling density.
Prepare SAQP.
See flow chart no.2 for more information on stratified, grid and judgmental sampling.

Prepare SAQP for sampling in accordance with Table 2, Section 5.2.5.

1. Fill of unknown origin can be indicated by the presence of buildings/structures, retaining walls on the site or at boundary. In addition, changes in ground levels when compared with the surrounding land and the presence of dam/pond walls also indicate the presence of fill. Fill suspected to have been sourced from site must still be investigated in accordance with Table 2. Section 5.2.5.

2. Sampling strategy will depend on site history. Site might need to be stratified with some sub-portions subjected to judgmental sampling and other areas to grid sampling.

Source: JBS&G.

Flow Chart 2: Choosing a sampling regime when there have been potentially contaminating activities

The site history has shown that potentially contaminating activities have been conducted at the site. Are there sub-portions of this site where there have been potentially contaminating activities that should be subject to judgmental (targeted) sampling? (See Section 5.1)

YES

NO

Stratify (divide) site into areas for judgmental sampling[1] and areas for systematic sampling.

Judgmental sampling
Develop SAQP in accordance with Sections 5.2.1.

Systematic sampling
Select a grid size based on the hotspot size you wish to detect. (Section 6). Then develop a systematic grid (section 5.2.2). Minimum sampling density cannot be less than that shown in Table 2 (section 5.2.5).
OR
Use minimum sampling density shown in Table 2 (Section 5.2.5). Develop SAQP.

1. Areas for judgmental sampling include pipelines, USTs, former/current structures, known leaks, identified ACM from site inspection etc.

Source: JBS&G.

Systematic sampling: density and justifications

Sampling density refers to the number of samples (n) per unit area (m^2 or ha) and can be written as n/m^2 or n/ha. For systematic sampling, the sampling locations must be spread over the site in a regular grid pattern, and the report should state what pattern has been used, for example, a square, offset square, random, triangular or herringbone grid pattern.

Table 2 in Section 5.2.5 shows the minimum sampling density for different site areas if the site history is incomplete, or if there is fill of unknown origin on the site.

A contaminated land report may contain a lesser or greater sampling density than that provided in Table 2, Section 5.2.5.

For a sampling density less than in Table 2

If the sampling density used is less than the minimum number of sampling points shown in Table 2, Section 5.2.5, the report should include:

- a justification for the density used that is based on a complete site history (see checklist for site history in EPA 2020b) – the site history must indicate that there have been no contaminating activities
- confirmation that a site inspection has been undertaken, and the inspection indicated that there was no evidence of contaminating activities, and no known or suspected fill of unknown origin on-site.

For a sampling density more than in Table 2

The sampling density can be greater than the minimum number of sampling points shown in Table 2, Section 5.2.5, if the site history and site inspection indicate that potentially contaminating activities have been carried out at the site.

Sampling density determined by hotspot size

Consultants can choose a sampling density based on a particular sized hotspot they wish to detect (Section 6, Part 1). A rationale for the hotspot size must be included and be based on the site history and observations made during a site inspection. Consultants can also include considerations of the hotspot shape in their calculations.

When number of samples required is determined by statistical methods

If the results of some sampling and analysis for the site are known, and the samples have been collected systematically from across the site and are considered representative of the current site conditions, then these results can be used to determine how many samples are required for site assessment. This can be done using the combined risk value method or the maximum probable error method (Section 7). These methods are not based on site area. These methods can also be used for calculating the number of samples required to assess a stockpile.

Worked example – stratified sampling regime

A site history and inspection have been completed for the site shown in Figure 17.

Table 11 lists each of the identified site sub-areas, the sampling regime proposed for each site sub-area, and the rationale for the chosen sampling regime.

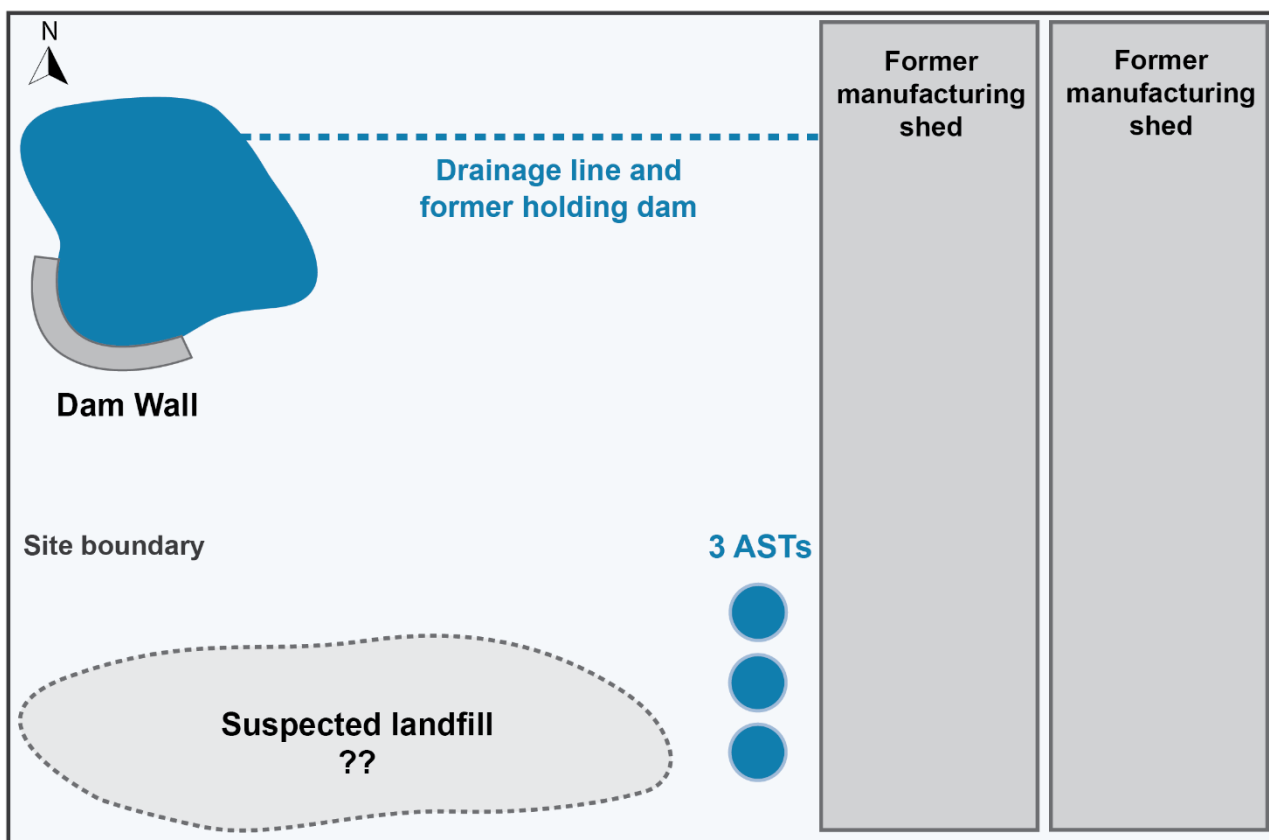**Figure 17   A site proposed to be subject to a stratified sampling regime.**

Table 11   Site sub-areas, sampling regime and the rationale.

Site sub-areas	Sampling regime	Rationale
Former manufacturing sheds and surroundings	Systematic. Potential judgmental (targeted)	Sheds have been demolished, but evidence of former retaining walls are visible, indicating fill of unknown origin. Building rubble is evident. Given the age of the former sheds (aerial views showed they were constructed in the 1930s) asbestos could be present, including friable asbestos from roofing material, in the soil surrounding the footprint of the sheds. Systematic sampling should be undertaken to characterise the fill material underlying the sheds. Field observations, laboratory results and statistical analysis of the laboratory results can all be used to characterise the fill material. Judgmental (targeted) sampling might be conducted if the inspection indicates it is required due to, for example, fill points for UST or staining/odours
Above ground storage tanks (ASTs)	Judgmental (targeted)	Inspection under the ASTs shows there have been leaks into the underlying soil. Sampling and analysis will help determine if remediation is required. Sampling should be directly underneath the ASTs. Field observations such as stained soil can guide sampling locations.
Suspected landfill (unknown extent)	Judgmental (targeted) at boundaries of landfill Systematic (within the landfill)	The boundary of the landfill is unknown, so targeted sampling locations will need to delineate the extent of the landfill. The nature of the material in the landfill needs to be assessed to determine if it is suitable to remain on site or will need to be

		classified for offsite disposal. This assessment should be made through systematic sampling.
Dam wall	Systematic	The dam wall is constructed from fill of unknown origin, and using a regular grid pattern is the most appropriate method for this type of material.
Drainage line and base of the former holding dam	Systematic	Boundaries of the drainage line and former dam are clearly evident and can be systematically assessed to ensure that sufficient samples are collected to be representative of the material encountered.
Remainder of the site	Systematic	Given the former industrial activities and the ground disturbance that has occurred at the site, the rest of the site should be subject to systematic sampling.

The following must be included in the report:

- site plans showing the sampling locations

analytical results tables clearly describing where samples were collected, with samples from each sub-area being presented together – for example, from the example in

- Table 11 all the dam wall results would be presented together in one table, but would be presented separately from the drainage line and the AST results.

If statistical analysis is required on the results from a site sub-area, only results from that sub-area should be included **and** the statistical analysis can only be used if the sampling was performed systematically. For example, if statistical analysis is required on the dam wall material, then only dam wall results should be included in that analysis.

The 95% upper confidence limit (95% UCL) should be presented for the analytical results for all analytes, if samples were collected systematically. NEPC 2013requires that:

- the 95% UCL is lower than the assessment criteria for all contaminants, AND
- no single value should exceed 250% of the relevant assessment criteria, AND
- the standard deviation of the results should be less than 50% of the assessment criteria.

The report should confirm that the NEPC 2013 requirements have been met and indicate how the 95% UCL and the standard deviation were calculated, including the name of any statistical software package used.

For more complicated data analysis, non-technical reviewers can refer to parts 1 and 2 of this document or require an audit under the *Contaminated Land Management Act 1997*.

Appendix I: Assessment process for stockpiles not impacted by asbestos

Flow Chart 3: Stockpile Assessment

Consider stockpile sampling principles (5.4.1)

Determine the purpose for stockpile sampling (5.4.2)

Onsite reuse (5.4.3)

Offsite reuse (5.4.4)
- Resource recovery
- VENM

Disposal to landfill or transport to recycling facility (5.4.5)

Onsite reuse (5.4.3)

Sampling should consider Table 3 and Table 4 in 5.4.6. Section 7 may be used in conjunction with Table 3 and 4

All samples, or the 95% upper confidence limit (95% UCL)[1] for all samples must comply with the assessment criteria[2]

- Complies with assessment criteria
- Does not comply

Stockpile may be reused onsite

Consider segregation or disposal to landfill or recycling facility, or whether orders and exemptions are applicable, or whether material can be kept onsite

---

Offsite reuse (5.4.4)

Resource recovery: Sampling must comply with relevant RRO and RRE

VENM: Sampling VENM may use Table 3 or Table 4

Sampling results must comply with relevant RRO and RRE, including any other conditions specified in the RRO and RRE
- Complies with RRO RRE
- Does not comply

Stockpile may be reused offsite

Sampling results must meet the definition of VENM. Stockpile must comply with the definition of VENM in the POEO Act
- Complies with VENM
- Does not comply

Stockpile may be reused offsite

Consider segregation but only if it is specified in the order and undertaken in accordance with the order, or disposal to landfill or recycling facility, or whether other orders and exemptions are applicable, or whether material can be kept onsite

---

Disposal to landfill or transport to recycling facility (5.4.5)

Sampling must comply with Table 3 or Table 4

Disposal facility: Sampling results must be assessed against Table 1 or Table 2 of the Waste classification guidelines, part 1; classifying waste, EPA 2014 ('the Guideline"), including the other steps of the Guideline to determine its waste classification. Waste must be taken to a lawful disposal facility able to receive the waste classification and waste type

Recycling facility: Sampling results must comply with the facility's Environment Protection License (EPL) conditions for receiving waste at the facility for recycling
- Complies with EPL
- Does not comply

Stockpile may be taken to recycling facility

Waste classification may be too high, e.g. Restricted Solid Waste (RSW), or Hazardous Waste (HW) present

Consider segregation or disposal to landfill if General Solid Waste or RSW, or processed at HW facility if HW, or whether material can be kept onsite

---

1. The NEPM states (Schedule B1, Section 3.2.1.) "At the very least, the maximum and the 95% UCL of the arithmetic mean contaminant concentration should be compared to the relevant Tier 1 screening criteria. However, where there is sufficient data available, and it is appropriate for the exposure being evaluated, the arithmetic mean (or geometric mean in cases where the data is log normally distributed) should also be compared to the relevant Tier 1 investigation or screening level".
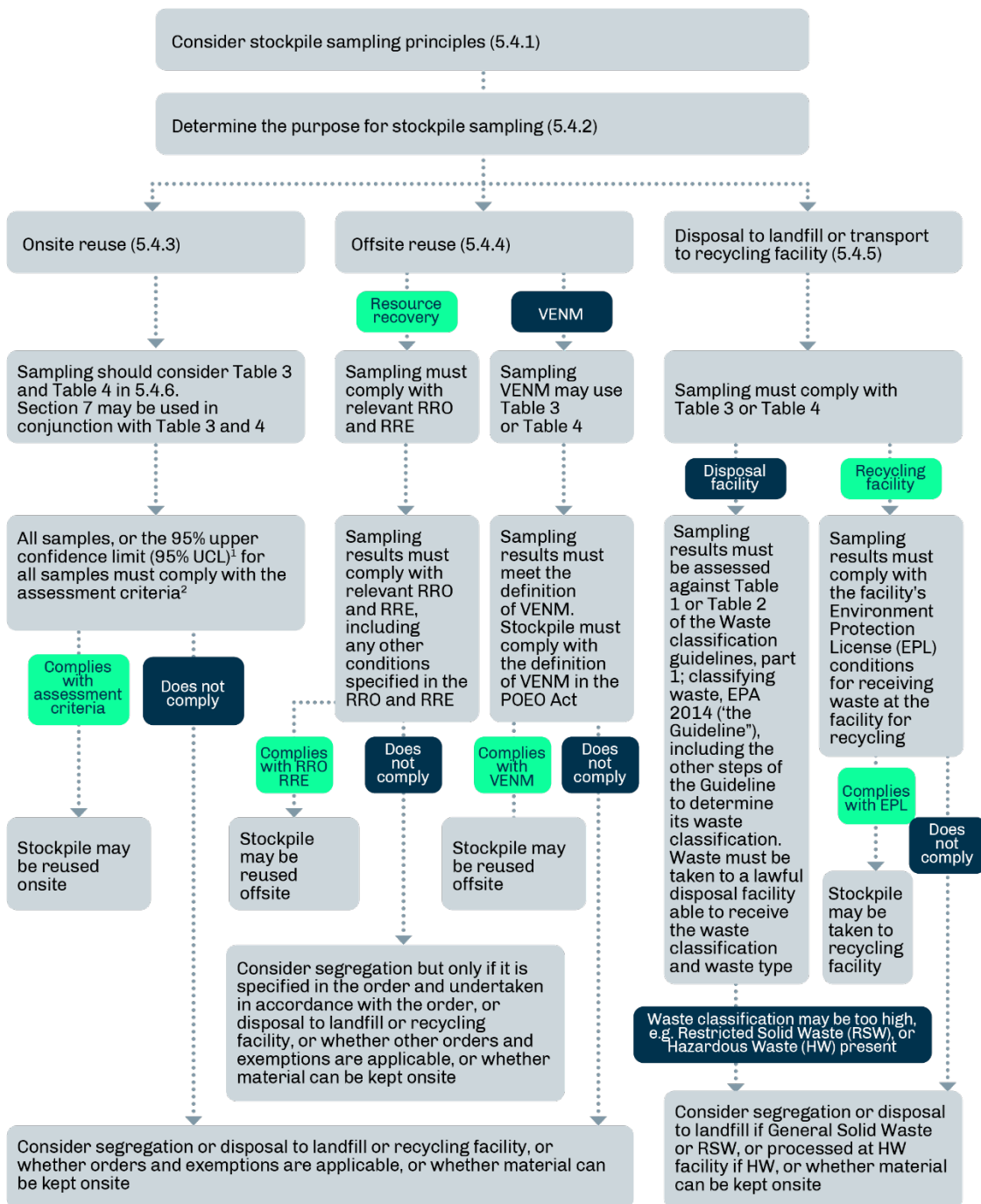
2. The NEPM (2013)[2] states that:
   a. The 95% UCL is compared with the assessment criteria for all contaminants, AND
   b. No single value should exceed 250% of the relevant assessment criteria. AND
   c. The standard deviation of the results should be less than 50% of the assessment criteria.

Sampling design part 2 - interpretation

Contaminated Land Guidelines

Acknowledgements

The NSW Environment Protection Authority acknowledges the contributions of the following organisations in the preparation of these guidelines:

Easterly Point Environmental Pty Ltd, JBS&G Australia Pty Ltd, Enviroview Pty Ltd, GHD Pty Ltd, Envirocene Pty Ltd, WSP Australia Pty Ltd, JK Environments Pty Ltd.

Cover: Prepared sample bottles and jars, sealed with EPA legal tape. Photo: EPA.

Contents

Figures

Tables

1. Introduction

The NSW Environment Protection Authority (EPA) has prepared these guidelines to assist contaminated-land consultants, site auditors, regulators, landholders and developers, and inform members of the public who have an interest in the outcomes of the assessment and management of contaminated land. They will help consultants to design sampling for contaminated sites, regarding where samples are collected, how many samples are collected, and ways the data is compared to relevant criteria.

The guidelines will help users obtain data that is appropriately representative for the purposes of the sampling and the media being sampled, and analyse and interpret the collected data.

As when following any guidance, users should justify the approaches they use, and demonstrate that they are appropriate and fit for purpose.

The guidelines are in two parts. The first part describes the application of sampling design; the second part (this document) provides guidance on interpreting the results. This second part is not a stand-alone document and should be read in conjunction with *Sampling design part 1 – application*.

These guidelines have been made in accordance with the *Contaminated Land Management Act 1997* (CLM Act). They should be read in conjunction with the CLM Act, the Contaminated Land Management Regulation 2013 (CLM Regulation), and any guidelines made or approved by the EPA under the CLM Act.

The guidelines complement other guidelines made by the EPA, and several national guidance documents that have been approved by the EPA. Those guideline documents are listed in References and cited in the text where appropriate.

1.1. Scope of these interpretation guidelines

Section 2

Information on comparing sampling data to action levels. Appendix A includes a summary of common descriptive statistics, and Appendices F to L show associated procedures and worked examples.

Section 3

Summary of the main statistical distributions and information on associated data transformations and data analysis.

Section 4

Introduction of the concepts of hypothesis testing, including decision errors and methods for conducting hypothesis tests. Procedures for common methods of hypothesis testing, along with worked examples, are shown in Appendices F to L.

Section 5

Information on confidence intervals for use in estimation problems, along with the use of upper confidence limits of the mean (UCL$\overline{x}$s) as another means of hypothesis testing. Appendices I to L provide procedures and worked examples for use of confidence intervals and upper confidence limits (UCLs), based on common distributions.

Section 6

Discusses trend analysis for temporal series of site contamination assessment data, including use of linear regression and the Mann–Kendall statistic.

Section 7

Discusses drawing conclusions from the data as part of the Data Quality Assessment process.

Section 8

Includes abbreviations and a glossary of technical terms.

Section 9

References for guidance and technical documents used in these guidelines.

1.2. Environmental media

These guidelines address sampling soil and solid media, as these are the most common targets when assessing site contamination. Information is also provided for other media, including groundwater, surface water, sediments and air.

Some statistical procedures described in these guidelines can be applied to these other media, although the EPA recommends that the following references are consulted when designing sampling programs for other media:

- NEPC 2013 – soil, groundwater and soil vapour
- ANZG 2018 – surface water
- DEC 2007 – groundwater
- EPA 2020a, DECCW 2010 and CRC Care 2013 – soil vapour
- Simpson and Batley 2016 – sediments.

This document does not specifically address biota sampling and ecotoxicity testing. For these areas, see the following references:

- ANZG 2018
- DES Queensland 2018
- DEC 2004.

2. Comparing data results to action levels

Schedule B1 of the *National Environment Protection (Assessment of Site Contamination) Measure 1999* (NEPC 2013) discusses the application of investigation and screening levels for Tier 1 assessments for soil results.

2.1. Use of statistics in the assessment of site contamination

Statistics can be broadly categorised as either **descriptive** statistics, which describe the sample, or **inferential** statistics, which relate the sample information to characteristics of the population. When assessing site contamination, both descriptive and inferential statistics are used to characterise sites and decision areas.

Descriptive statistics are discussed further in Appendix A. See the Glossary for more definitions of statistical concepts.

For **inferential statistics**, tests can be **parametric** or **non-parametric**. Parametric statistical tests make assumptions about the parameters of the population distribution, whereas non-parametric tests (sometimes called distribution-free statistics) make no assumptions about the distribution although they may make assumptions about the data.

All parametric statistical tests assume that the data are drawn from a particular probability distribution – normal, log-normal, gamma, or some other known statistical model. Parametric tests generally have non-parametric counterparts which can be used when the assumptions of the parametric test cannot be met. As non-parametric tests do not make assumptions about the distribution, they typically have lower statistical power than parametric tests in cases where the assumptions hold. However, non-parametric tests are often more accurate and powerful than parametric tests for even modest departures from parametric test assumptions.

Two assumptions that apply to many forms of inferential statistics are, first, that the sampling data are unbiased and, second, that each member of the population has an equal chance of being included in a sample. Consequently, the data points are an independent and identically distributed sequence of observations. **Independent** means that each observation is not controlled by the value of any other observation. Independence can generally be assumed for random samples if the sample consists of less than 10% of the population. **Identically distributed** means that the samples have been taken from a parent population whose mean and variance is stationary over the space and time of collection.

Biased sampling can be judgmental (also known as targeted) and arbitrary where certain observations are included or excluded because of some feature: this leads to members of the population having an unequal opportunity of being sampled. Bias can arise from a subconscious decision by the sampler.

Basic statistical tests can be validly applied only to unbiased sample data; data from judgmental or arbitrary sampling should not be used for statistical tests. For this reason, it is recommended that data obtained using a combination of judgmental and random (probabilistic) sampling approaches is collated and considered separately, and that the formal use of statistical techniques is confined to probabilistic sample data only. This means that results from judgmental sampling – for example, validation of an excavation, or investigation of a contaminating feature such as a leaking pipeline – should be removed from a dataset before statistical analysis is performed on the remaining data.

2.2. Descriptive statistics

Terms used in statistics are referred to as **statistical descriptors**. Commonly used statistical descriptors are the sample **range**, sample measures of central tendency (**mean**, **median** and **mode**), sample **percentiles**, and sample variability (**variance**, **standard deviation** and **coefficient of variation**). A preliminary data review could include basic graphical representations of the data, such as spatial plots,

box and whisker plots, frequency plots, histograms, ranked data plots, quantile–quantile (Q–Q) plots, two variable scatterplots, and temporal plots[1].

Common statistical descriptors can be used to summarise the basic quantitative characteristics of the sampling data, allowing them to be presented in tables or illustrated graphically. Where there are multiple populations or decision areas, it is useful to separate the data for analysis and comparison to reduce the variability of the individual datasets.

Reviewing the data numerically and graphically leads to a better understanding of the structure of the data, reveals patterns in distribution and relationships, and identifies potential anomalies. Data should be verified and validated before it is reviewed.

Descriptive statistics are further summarised in Appendix A, and specific procedures for determination and worked examples are included in Appendices B to D.

2.2.1. Software tools and packages

Statistical software tools and packages are available in spreadsheets, commercial software and open-access freeware. These can be used to determine both descriptive and inferential statistics. Freeware is particularly recommended as it allows other stakeholders including auditors and regulators easy access for checking outputs, without the associated financial costs and licence restrictions associated with commercial products.

A detailed review and summary of widely available statistical software packages can be found in Appendix D of ITRC 2013. This review covers both general statistical packages for broad applications and packages specifically designed for statistical analysis of environmental data, and includes both commercial and open-source freeware.

2.2.2. Data presentation

Spreadsheets and statistical software tools and packages can create sophisticated outputs to represent the sampling data and associated statistical information. For a preliminary data review, for example, they can present data in plan and cross section, both spatially and temporally, and as graphics.

As noted by DoE 1998:

> *While reporting of minima, maxima, mean, median, standard deviation, upper confidence limits etc. provides necessary information, such data may not be sufficient to characterise a site. The use of histograms or frequency distributions should also be considered to illustrate the distribution of results.*

Appendix E gives examples of graphical presentations that can be easily developed.

2.3. Maximums

The maximum observed value in a dataset is important in assessing site contamination, as a site or decision area is generally considered suitable for the intended land use if the maximum observed value is below the criterion or action level. However, such a condition may be misleading. The maximum observed value of the contaminant of interest is unlikely to be the maximum value present in the population, and the relationship between the two cannot be determined in the absence of statistical analysis.

Where sampling data is highly variable or based on small sample sizes, it may not be representative of the underlying population's variability and decision errors can arise. The recommended approach to control decision errors is to conduct appropriate tests that allow statistical inference. Appropriate tests include hypothesis tests, such as one-sample t-tests and UCL$\bar{x}$s. Section 4 and Appendix F discuss hypothesis testing; Section 5 and Appendices I to L for skewed distributions discuss UCLs.

---

[1] See Sections 14.3–14.6 of Schedule B2 of NEPC 2013, USEPA 2006a G-9S and USEPA 2006c G-9R for more details.

When comparing sample results to criteria and action levels, the sampling data needs to show that no single value exceeds 250% of the relevant investigation or screening level (schedule B1, NEPC 2013).

2.4. Outliers

In statistics, **outliers** are data points that do not fall into the expected range of a defined probability distribution function. In the context of site contamination assessment however, the characteristics of a probability distribution function of a contaminant can be difficult to define. Complex historical site uses can result in the superposition of multiple probability distribution functions. **Hotspots** – small areas of relatively high concentration – may also be present, with their own probability distribution function.

Discarding an outlier from a dataset should be done with extreme caution as environmental datasets often include legitimate extreme values (USEPA 2006b). A more thorough examination of the reasons for any unexpectedly high values may lead to new insights into the data such as the presence of an unsuspected hotspot of contamination, or to reconsidering underlying assumptions about the data and its distribution.

All data resulting from probability-based sampling must be included in the subsequent inferential statistical analysis, unless:

- it can be demonstrated with a high level of confidence that the individual data points are invalid due to transcription errors, data coding errors, or measurement errors in the laboratory analysis

    or

- the individual data points are subsequently identified – again, with a high level of confidence – as part of a hotspot, and the hotspot is appropriately remediated or managed and thereby effectively removed from the population.

In either case, a determination is then needed as to whether further data needs to be generated through additional investigations, or if sufficient data is available to support the required decisions. These determinations should include appropriate statistical analysis of the remaining dataset.

2.5. Non-detects

As part of the assessment of site contamination, where the concentration of an analyte ranges between zero and the limits of reporting (LORs) of the laboratory method, the results are reported as less than the LORs. This is referred to as **left-censored** data. In some instances, the data below the LORs may represent another **population**, and the data, including geological logs and field notes, should be reviewed to determine if a more appropriate grouping of data is relevant. For determining mean values, mixing a large number of results below the LORs with a limited number of detected results can lead to estimation problems if simplistic methods are used.

There are various imputation methods to replace these censored values. **Direct substitution** is the easiest but least satisfactory. Generally, substitution should only be adopted where the fraction of the sample that is censored is relatively small. With substitution, a constant value is assigned to the non-detects by one of the following:

- assuming the non-detects are equal to zero
- assuming the non-detects are equal to the LORs
- assuming the non-detects are equal to some fraction of the LOR, usually one half.

The proxy value is then used as though it were the value for that measurement. However, the uncertainty associated with the substitution method increases as the proportion of non-detects in the dataset increases. Statistical determinations and inferences associated with censored data become increasingly problematic, because of errors in the estimates of parameters such as the mean, which becomes biased down. The direction and extent of the bias in variance is highly dependent on the data and substituted value.

Some statistical software packages have methods that enable a user to enter data and indicate that it is a non-detect; the software calculates statistical parameters such as 95% UCL and standard deviations for the dataset, and the output of the statistical package provides guidance on which method is recommended (USEPA 2015a). A worked example is provided in Appendix M.

If statistical software is unavailable, Section 4.7 of USEPA 2006a provides more detailed guidance for analysing data with non-detects. If the direct substitution method described above is used, results for the three substitutions listed above – zero, equal to LOR and assumed fraction of LOR – should be reported.

Where non-detects below LORs exist::

- always report detection limits for non-detects
- do not convert non-detects to zeros without specific justification
- consider using non-parametric methods if further statistical analysis is required.

Other methods of imputation – replacing data with substituted values – include **multiple imputation**, **fractional imputation** and **Bayesian modelling**. The appropriate imputation method depends on the size of the dataset and the proportion of measurements reported as non-detects. If the proportion of non-detects is high (> 50%) or the number of samples is small (n < 5), analysis may be challenging.

The method of **maximum likelihood** by first principles can be used to estimate the parameters of a probability distribution even where there is censored data: for censored data points, summing is replaced by integration between limits (zero and LOR). In general, the point of maximum likelihood cannot be determined algebraically but must be solved numerically (for instance, with the hill climbing technique or Newton Rapson technique), though this is no longer an issue for those with access to desktop computing.

Additionally, there are various statistical packages dealing with censored data that are suitable for laboratory measurements.

Refer to ITRC 2013 and USEPA 2015a for details of specific methods for managing non-detects in statistical analyses.

Wendelberger and Campbell 1994 note that:

> [t]he manner in which the nondetect values are handled should depend on the type of decision to be made and the magnitude and frequency of the nondetect values. If the nondetects are small in magnitude or low in frequency, the method of handling the nondetects will probably have minimal impact on the final outcome of the analysis. However, if the detection limits are close to important decision values, or if the frequency of nondetects is high, the treatment of the nondetect values can greatly influence resulting decisions.

Whichever statistical approach is adopted, the conceptual site model (CSM) should be re-developed to reflect the proportion of the dataset samples that are non-detects. For instance, if a site has a few detections and many non-detections, the source of the contamination should be carefully considered when refining the CSM. An option might be to stratify the site so areas where there are widespread non-detections are assessed separately from areas with detections, especially if investigation levels are being exceeded.

2.6. Pseudoreplication

In site contamination assessment, the collection and analysis of duplicate and triplicate samples is conducted as part of quality assurance/quality control (QA/QC) programs. Although this is important for determining the data's usability, these replicate sample results must not be treated as an independent sample. Doing so is known as **pseudoreplication** because the duplicates and triplicates are not independent of the primary sample. Pseudoreplication increases the number of samples while providing another data point similar to the primary sample, resulting in bias and distortion of any statistical analysis being undertaken.

2.7. Contaminant distribution

The variation of contaminant concentrations over a site or decision area means that individual measurements cannot be used to fully describe the distribution of a contaminant. If the contaminant concentrations are plotted against their respective frequency of occurrence, the resulting curve or histogram represents the concentration distribution of that contaminant over the site or decision area.

While histograms inform the characterisation of the site or decision area, they do not represent spatial information across the site or decision area. They show the range, central tendency, variation and

distribution of the variables being considered. Under the **multiple lines of evidence and weight of evidence** approach, these parameters should be considered when interpreting the data and comparing it to the criteria or action levels.

For example, comparing the sampling results to 250% of the relevant investigation or screening level can lead to identifying apparent hotspots and the recommendation that removing these areas will make the site or decision area suitable for the proposed land use. However, closer examination of the data may show that the apparent hotspots relate to heterogeneity of the soil or fill and any subsequent validation would result in identifying other 'hotspots'. In these situations, more characterisation to confirm the variability of the soil or fill may provide better information for decision making on remediation or management. In such situations, the use of statistical tools can assist, particularly in relation to decision errors and determining a suitable number of samples.

Schedule B1 of NEPC 2013 requires that sampling results should be checked to ensure the standard deviation of the variable is less than 50% of the relevant investigation or screening level. Although 50% is an arbitrary value, it provides a warning that the variance is potentially excessive, prompting further review of the contaminant distribution.

In these cases, further segregating the data by, for example, depth, soil type or spatial distribution may demonstrate that multiple populations are inappropriately being considered as a single population. Alternatively, the data may indeed represent a highly variable population and indicate more sampling is needed.

3. Distributions, transformations and data analysis

The sampling distribution is the frequency or probability of occurrence of measured values. In the assessment of site contamination, data can be analysed using parametric (distribution based) methods, or non-parametric methods where the population is not assumed to fit a specific population distribution. Statistical software packages provide more complex calculations of UCL$\bar{x}$ using a number of parametric and non-parametric distributions, however a brief review of the predominant distributions used is warranted.

Where the sampling data has a normal (or more strictly, nearly normal), log-normal or gamma distribution, parametric methods can be applied. Where the sampling data does not have such a distribution, non-parametric methods should be used. Non-normal datasets can have a transform applied to essentially normalise the data, which aids analysis.

3.1. Parametric methods

Population parameters are estimated from samples. Different random samples will produce different estimates of each parameter; for instance, each sample will produce a different estimate (using $\bar{x}$) of the population mean, μ. These estimates themselves have a distribution, known as the **sampling distribution**. Many common statistical methods are based on a knowledge of, or the assumed characteristics of, the sampling distributions of population parameter estimates.

3.1.1. Normal distribution

The most commonly used distribution in parametric statistics is the normal. The central limit theorem (CLT) states that the sampling distribution of the mean for **n** independent random samples approaches a normal distribution as **n** increases. This holds for all population distributions with finite mean and variance. With normal distribution, the mean, median and mode are equal.

Based on the CLT, the sampling distribution of $\bar{x}$ can be approximated by a normal distribution when the sample size n is sufficiently large (> 30), irrespective of the shape of the population distribution. The larger the value of n, the better the approximation (Devore and Farnum 2005).

Appendix J demonstrates the one-sided Student's t-test method used with normal distribution.

3.1.2. Log-normal distribution

Log transformations convert samples to natural log values to allow the use of log-normal or exponential distributions for analysis. The log-normal is a continuous distribution in which the logarithm of a variable has a normal distribution. Thus, if the random variable x is log-normally distributed, then y = ln(x) has a normal distribution. Likewise, if y has a normal distribution, then the exponential function of y (that is, x = exp(y)) has a log-normal distribution.

In log-normal distributions, the mean, median and mode are not equal. The difference between mean and mode depends on the skewness of the population, while the median is independent of skewness.

Appendix K gives a worked example of the Land's H-statistic method used with log-normal distribution.

3.1.3. Gamma distribution

Gamma distribution is more flexible for fitting data than normal and log-normal distribution. Gamma distribution is a rank-order transformation where contaminant concentration data is sorted into ascending order and converted to an integer ranked list. This process eliminates the scale effects in contaminant concentrations commonly found in site contamination datasets and reduces the effect of large differences between results in a dataset.

This distribution type is relevant to the assessment of contaminated sites due to its relationship to exponential and normal distributions. The **gamma distribution** is a two-parameter family of continuous

probability distributions. The exponential distribution and the chi-squared distribution are special cases of the gamma distribution.

The three common parametrisations for gamma distributions are:

- a shape parameter k and a scale parameter θ
- a shape parameter α = k and an inverse scale parameter β = 1/θ, called a rate parameter
- a shape parameter k and a mean parameter μ = kθ = α/β.

In each of these, both parameters are positive real numbers and control the shape and skewness of the distribution.

3.1.4. Parametric methods in the analysis of site contamination assessment data

When using distributions to assess site contamination data, the limitations imposed by each distribution must be accounted for as these determine how well the distribution can provide a reliable interpretation of the actual population.

Data for assessing contaminated sites is rarely normally distributed, due to the kind of processes that lead to site contamination. When the mean, median and mode are not equal, or the coefficient of variance is > 0.5, consider carefully before using the normal distribution for analysis. Similarly, be cautious when applying the log-normal distribution, as the data for assessing contaminated sites is often not truly log-normally distributed. The application of either distribution needs to the be verified by testing that the data is approximately normally distributed, or normally distributed after the log transform is applied. The distribution can be tested by using a statistical software package to construct quantile–quantile (Q–Q) plots, which graph the quantiles of the dataset against the quantiles of a specific probability distribution.

It is generally recommended that skewed datasets are assessed using a gamma distribution rather than a log-normal distribution, as this produces more reliable results. A log-normal transformation disguises the effect of high values that may not represent background and exaggerates the apparent standard deviation of the modelled log-normal distribution. This increases the risk of making an incorrect decision in relation the population distribution and associated statistical parameters. Therefore, for assessing skewed site contamination datasets, the gamma distribution should be used when performing parametric analysis, particularly if the sample size is less than 20 and/or contains outliers. Because of the gamma function's flexibility in accommodating a wide range of symmetric and asymmetric (skewed) distributions, it can represent log-normally distributed datasets without the risk of masking the effects of outliers.

When the site assessment data is highly skewed by extreme values or a significant number of non-detect values, it may be hard to determine an appropriate distribution for parametric analysis. In such cases non-parametric methods may give more reliable results.

3.2. Non-parametric methods

Non-parametric statistics are analysis methods that either make no assumption about the distribution of the data or the population, or, where a specific distribution is assumed, do not specify the distribution's parameters. Commonly used non-parametric methods for making inferences in the assessment of site contamination data are the **bootstrap**, **jackknife** and **Chebyshev** methods.

Compared with non-parametric methods, analogous parametric methods are usually more effective when the assumptions of parametric methods hold. Where a population departs from these assumptions, the non-parametric tests can be superior.

3.2.1. Bootstrapping

Bootstrapping involves estimating properties of a statistical parameter by measuring those properties through randomly re-sampling the dataset with replacement data. Data points need to be independently and identically distributed. This 'new' dataset is then used to estimate the statistical parameters such as mean, median, mode and standard deviation. Bootstrapping can also be used for constructing hypothesis tests, as an alternative to statistical inference based on the assumption of a parametric model, when that assumption is in doubt.

Bootstrapping, like any non-parametric resampling method, offers a useful means of reducing the influence of extreme outliers on the overall statistical parameters of the underlying sampled population. However, caution is needed to avoid diminishing the importance of outliers in relation to the overall decision: where the outlier represents a hotspot, a non-parametric re-sampling method such as bootstrapping may not be appropriate. The use of this method must therefore be justified in the context of the importance of the outliers to the overall decision being made.

3.2.2. Jackknifing

Jackknifing is similar to bootstrapping, in that the method re-samples the dataset and generally produces similar results, although instead of making random replacements, it randomly removes a sample in each resampling step. The re-sampled dataset can then be analysed with the same methods as those used for bootstrapping. Jackknifing is subject to the same limitations and cautions as bootstrapping.

3.2.3. Chebyshev

The Chebyshev method is a non-parametric method that does not involve resampling the dataset but instead relies on use of the Chebyshev's inequality. This specifies that, for all distributions with finite mean and variance, only a certain fraction of values can be more than a certain distance from the mean, that is, no more than $1/k^2$ of the distribution's values can be more than k standard deviations away from the mean with 100% certainty.

This inequality can be applied to any probability distribution in which the mean and variance are defined. When applied to datasets for assessing site contamination, the Chebyshev method can determine statistical parameters, particularly the mean, for highly skewed datasets or ones that contain significant outliers. In most applications, the Chebyshev method gives a more conservative result than other parametric and non-parametric methods.

Appendix L gives a worked example of the Chebyshev method.

4. Hypothesis testing

Decision problems can be addressed as statistical hypothesis tests, which are recommended under the USEPA's DQOs (data quality objectives) process. The **null hypothesis significance testing** (NHST) framework, derived from approaches for testing data, is a method of statistical inference used to determine if a null hypothesis (H_0) should be rejected in favour of an alternative hypothesis (H_A) at a specified level of confidence. In the assessment of site contamination, H_0 is that the site or decision area is not suitable for the specified use, i.e. that the site or decision area is contaminated.

The basis of the hypothesis test is that H_0 can be rejected where the findings are incompatible with H_0 being true, in which case H_A is more likely. Alternatively, H_0 can fail to be rejected, which does not necessarily mean that H_0 is true but that there is not enough evidence to reject the site being contaminated.

Before testing, an environmentally significant difference from the criterion level should be established. For H_0 to be rejected, the data must show, with given confidence, that the population parameter is at or below this level. This environmentally significant difference is greater or equal to zero to provide an environmental buffer.

The most common form of hypothesis testing is for nearly-normally distributed populations, where estimated population means are tested using the Student's t test (t-test) which is used to test for differences in population means. This test can be:

- a **one-sample t-test**, to test whether the mean of a single population is different from a target value, such as a specified health investigation level (HIL)
- a **two-sample t-test**, to compare the means of two groups, such as site data and background data
- a **paired sample t-test**, to compare the means from the same group at different times, such as before and after remediation.

If there are non-detects, more work is required to estimate the mean and variance.

Worked examples are shown in Appendix F (a one-sample t-test) and Appendix G (a two-sample t-test).

There are also comparable parametric methods for non-normal distributions, and non-parametric methods for testing differences in means and/or medians in unknown distributions (see USEPA 2006a, G-9S).

4.1. Sampling uncertainty and decision errors

Uncertainty in estimates is unavoidable due to, for example, inherent variability in the characteristics of interest of the target population, the limits on the number of samples that can be collected and imperfect measurements. Statistical methods provide quantitative tools for characterising an estimate's uncertainty, and help in designing an investigation that will generate probabilistic data of a sufficient type, quality and quantity.

One can never be 'certain' about an answer derived from sampling, so the uncertainty must be specified for a statistical statement to have meaning. In statistics, uncertainty is technically referred to as **risk** or **confidence level**. The risk of incorrectly rejecting H_0 is denoted by α (alpha) and has a magnitude of between 0 and 1. The risk of incorrectly accepting H_0 is denoted by β (beta), which is also between 0 and 1. For example, if a particular statistical statement is quoted as having a 95% confidence level, ($\alpha = 0.05$), this implies that at least 95 out of I00 repeats of the sampling will correctly accept a true H_0. A power of 80% (β = 0.2) means an 80% chance of correctly rejecting a false H_0.

In the assessment of site contamination, α risk is the risk of deciding that the site or decision area is suitable for the proposed use when in fact it is not, and the confidence level is always equal to 1 - α. The probabilities generally used in the assessment of site contamination are $\alpha = 0.05$ and β = 0.2, or a 95% confidence level and a statistical power of 80%, although higher probabilities can be used, such as $\alpha = 0.01$ and β = 0.1, or a 99% confidence level and a statistical power of 90%.

Changing one probability inevitably changes the other. One way to obtain both a high confidence level and high statistical power is to increase the number of samples. More sophisticated sampling designs and associated analysis can also be used to increase the power – see Section 4.

Within hypothesis testing, **decision errors** refer to the incorrect decisions that can be made about a site or decision area, based on the data collected. They arise from using data that are not sufficiently representative of the site or decision area due to sampling errors, measurement errors or more commonly, both. Such errors can lead to decisions that assess contaminated land as uncontaminated when it is contaminated, or that determine that remediation is required when it is not. The combination of errors from all sources is referred to as the **total study error**, and directly affects the probability of making decision errors. The statistical theory behind hypothesis testing allows the probability of making a decision error to be quantified, given the data collected and the specified level of significance.

Decision errors result from:

- **sampling errors**, which arise from using information from a sample instead of measuring the whole population
- **sampling design errors**, which arise when the sampling design does not validly capture the structure of the population – they include sampling frame selection, sampling unit definition, selection probabilities and the number of samples collected
- **measurement errors**, which arise from the variability inherent in sample collection, handling, preparation, analysis and data reduction.

Study error is managed by correctly choosing suitable sampling designs and measurement systems.

See Appendix H for more information on the types of decision errors.

4.2. Use of hypothesis tests

Formal statistical methods can quantify the uncertainty associated with decisions. ITRC 2103 notes common decision errors when assessing site contamination, and hypothesis tests that can control them:

- **concluding that a site or decision area is suitable when the sample maximum is less than the criterion or action level.** For some distributions and sample sizes, the population mean of the site or decision area may be greater than the criterion or action level, even though a particular sample maximum is less than the criterion or action level.

  This is a Type I decision error, and a one-sample hypothesis test will allow statistical inference and control of decision errors.

- **concluding that a site or decision area is not suitable when the sample maximum is more than the criterion or action level.** The population mean of the site or decision area may be less than the criterion or action level when the sample maximum is more than the decision criterion, depending on the nature of the distribution and the sample size.

  This is a Type II decision error, and a one-sample hypothesis test will allow statistical inference and control of decision errors. The systematic planning for the investigation should describe how maximum values will be treated – for instance, what further data analysis or investigation will be carried out if the maximum value exceeds 250% of criterion or action level.

- **concluding that the failure to reject the null hypothesis 'proves' the null hypothesis (i.e. that the site is too contaminated to be acceptable).** As environmental data typically shows large random variability, the sample could include a preponderance of elevated concentrations, particularly if the sample size was small and so the statistical test is of insufficient power.

  The power of the test ($1 - \beta$) should be determined and compared to the decision criteria, and/or the number of samples required to achieve the specified decision criteria determined using the combined risk value (CRV) method discussed in *Sampling design part 1 – application*. If not enough samples were collected (i.e. the test was conducted with insufficient power) further data analysis or investigation may be required

- **directly comparing the maximum value of a site or decision area with a background maximum or mean, without considering potential decision errors.** The maxima from the two datasets should not be compared to make inferences about the means of the datasets, as decision errors are

not controlled for and this can result in Type I errors. A two-sample hypothesis test is recommended to allow statistical inferences and control decision errors.

5. Confidence intervals and upper confidence limits

A **confidence interval** estimates a population parameter from sample data and is composed of two parts: an interval calculated from the data and a confidence level associated with the interval. In the assessment of site contamination, the confidence interval is generally expressed as a point estimate, usually the mean plus and minus (±) the margin of error. Because confidence intervals are expressed in this way, they are determined using two-sided intervals for the t critical values.

Upper confidence limits (UCLs) are the upper component of the confidence interval and are determined using the one-sided interval for the t critical values.

Hypothesis tests and confidence intervals are related, as they are determined using variations of the same formula, and often a confidence interval can be used to test a hypothesis, making it unnecessary to perform the entire hypothesis test. In assessing site contamination, a decision is generally only required on whether the estimated population parameter exceeds the criterion or action level, so UCLs can be used by themselves as a form of hypothesis testing.

5.1. Confidence intervals

Performance criteria are needed to estimate an unknown parameter to within a specified amount, with a given confidence level: they specify the maximum width of the confidence interval. The width of a confidence interval depends on the number of samples used to calculate the interval, the precision or variability of the dataset and the specified confidence level. Placing limits on the maximum width of a confidence interval enables the precision and the number of samples needed to calculate it to be determined. As the variability of the population being studied is generally fixed, only the confidence level and number of samples can be controlled.

For independent samples from an approximately normal distribution, or where the sample size is large (n ≥ 30), confidence intervals for mean values are determined by using the one-sample Student's t-test. This test is reasonably robust if the population distribution deviates only moderately from normality; however, for highly skewed datasets with significant outliers, or where significant non-detects are included in the dataset, other distributions or non-parametric methods should be used.

Appendix F shows how to determine confidence intervals using the one-sided Student's t-test and gives a worked example. For other distributions not discussed below, or for non-parametric methods, see USEPA 2006a, G-9S.

5.2. Upper confidence limits

When assessing site contamination, the main way to determine if sites or decision areas are suitable for their proposed uses is to employ UCLs as one-sided hypothesis tests for comparing the sample mean to the action levels or criteria. The appropriate method is determined by the population distribution, as indicated by the sampling data.

- For normal distributions, Appendix J demonstrates the one-sided Student's t-test method.
- For log-normal distributions, Appendix K demonstrates the Land's H-statistic method.
- For skewed distributions, Appendix L demonstrates the Chebyshev inequality method.

6. Trend analysis

Trend analyses are used in the assessment of site contamination to determine if a contaminant's concentrations are increasing, decreasing or remaining constant over time. The objective of a trend analysis is to determine if the changes of a contaminant concentration can be statistically correlated to time and, if so, how significant the correlation is. The two trend analysis methods described below are generally applied to datasets for water or air, although they can be used for assessing remediation, such as the bioremediation of soil.

6.1. Linear regression

The calculation of a linear regression, or line of best fit, is a common way to measure the relationship between two variables. In the assessment of site contamination, a linear regression analysis is often used to assess if there is a trend between a contaminant concentration and time, for example, is the concentration of benzene in monitoring well two decreasing over time? The need for temporal trend analysis and the minimum number of data points required depends on the CSM, which should be developed based on site specific characteristics.

Data should be presented on a time plot to determine, visually, if a trend is likely, then the r-value (or Pearson Correlation Coefficient) should be calculated. This is a measure of the strength of the linear relationship between the two variables. The r-value can be a value between 1 and -1, with 1 indicating a strong positive relationship between the two variables, -1 indicating a strong negative relationship, and 0 indicating no relationship at all.

While the calculation of an r-value of 0.98 may indicate a strong positive relationship between the contaminant concentration and time, other factors could be affecting this relationship. Simple linear regressions can be affected by seasonality, the distribution of the data and the number of samples below the LORs. USEPA 2006a, G-9S states that due to these limitations, linear regressions are not generally recommended for estimating and detecting trends but can be used as an informal and quick screening tool to detect if a strong linear trend is present.

6.2. Mann–Kendall

The Mann–Kendall test is used to assess trends in datasets, and being a non-parametric test, it makes no assumption regarding data distribution and is unaffected by missing data or values below the LORs.

The test compares each data point against the next data point, and a score of 1 or -1 is given for each comparison, according to whether there is an increase or decrease in concentration. The test is not affected by the magnitude of the change.

The individual scores are tallied to provide the Mann–Kendall statistic (S): a positive S indicates an upward trend whilst a negative S indicates a downward trend. The value of S is then compared to an S-critical value. A p-value is then calculated for comparison with the adopted significance level, which determines if the null hypothesis (of no trend) is rejected or accepted.

The Mann–Kendall test is also affected by seasonality, and only data from similar months each year should be compared if this is likely to be important. Where high seasonality effects can be expected, to calculate a meaningful result means collecting data for at least four years.

The output of the Mann–Kendall test will be:

- the concentrations are increasing, or
- the concentrations are decreasing, or
- there is no trend.

Following this test, a linear regression analysis can be performed to determine the strength of the trend, providing the potential limitations of the linear regression are considered. Further information on the use of the Mann–Kendall test to assess trends can be found in Gilbert 1987, USEPA 2006a, G-9S and IRTC 2013.

7. Drawing conclusions from the data

Once the investigation has been conducted and validation data have been collected for a project, the consultant should consider if the data quality objectives have been met by referring to the data quality indicators determined during the DQO process.

The consultant should document any statistical calculations clearly, evaluate the results and draw conclusions. If warranted, the CSM for the site should be updated to incorporate any analytical data. These should be presented in an assessment report prepared in accordance with EPA 2020b.

Further details on drawing conclusions in accordance with the DQO process are provided in USEPA 2006a.

Guidance on assessment for media other than soil or fill is provided in:

- NEPC 2013 – soil, groundwater and soil vapour
- ANZG 2018 –surface water
- DEC 2007 – groundwater
- EPA 2020a, DECCW 2010 and CRC Care 2013 – soil vapour
- Simpson and Batley 2016 – sediments.

See EPA 2020b for advice on preparing reports.

8. Abbreviations and glossary

8.1. Acronyms and abbreviations

ABC	Ambient background concentration
ANZG	Australian and New Zealand water quality guidelines
CECs	Contaminants of emerging concern
CLM	Contaminated land management
CLT	Central limit theorem
CoPC	Contaminants of potential concern
CRV	Combined risk value
CSM	Conceptual site model
CV	Coefficient of variation
DNAPLs	Dense non-aqueous phase liquids
DQIs	Data quality indicators
DQOs	Data quality objectives
DSI	Detailed site investigation
DUs	Decision units
EPA	Environment Protection Authority
HIL	Health-based investigation level
HSL	Health screening level
ISM	Incremental sampling methods
LNAPLs	Light non-aqueous phase liquids
LOR	Limits of reporting
Metals	Arsenic (As), cadmium (Cd), chromium (Cr), copper (Cu), lead (Pb), mercury (Hg), nickel (Ni) and zinc (Zn)
MoE	Margin of error
MPE	Maximum probable error
MQOs	Measurement quality objectives
NEPM	National Environmental Protection Measure
NHST	Null-hypothesis significance testing
NOW	New South Wales Office of Water
OEH	New South Wales Office of Environment and Heritage
PAHs	Polycyclic aromatic hydrocarbons
PFAS	Per- and poly-fluorinated alkyl substances
PFHxS	Perfluorohexane sulfonate
PFOS	Perfluorooctane sulfonate
PFOA	Perfluorooctanoic acid
PSH	Phase-separated hydrocarbon
PSI	Preliminary site investigation

PID	Photoionisation detector
PTFE	Polytetrafluoroethylene
QAPP	Quality assurance project plan
QA/QC	Quality assurance/quality control
Q–Q	Quantile–quantile
RAP	Remediation action plan
RSD	Relative standard deviation
SAQP	Sampling and analysis quality plan
SOPs	Standard operating procedures
STP	Sewage treatment plant
SWL	Standing water level
TOFA	Total organic fluorine assay
TOPA	Total oxidisable precursor assay
TRHs	Total recoverable hydrocarbons, including volatile C6–C10 fractions and semi- and non-volatile C11–C40 fractions
UCLs	Upper confidence limits
UCLx̄	Upper confidence limits of means
UPSS	Underground petroleum storage system
USEPA	United States Environmental Protection Agency
UST	Underground storage tank
VOCs	Volatile organic compounds

8.2. Statistical notations

$1 - \alpha$	Confidence level
α	Type I error rate (see Glossary)
β	Type II error rate (see Glossary)
c	Criterion/action level
df	Degrees of freedom
exp	Exponential function
H_A	Alternative hypothesis
H_0	Null hypothesis
n	Number of samples or measurements in a sample (see sample definition)
θ	Scale parameter of the gamma distribution
σ	The population standard deviation, which is generally not known
σ^2	The population variance, which is generally not known
p-value	Probability value
Δ	Uppercase Greek letter delta, denoting the width of the grey region associated with hypothesis testing
s	The sample standard deviation, which is determined from the measurements taken
s^2	The sample variance, which is determined from the measurements taken
δ_0	Difference (delta) of zero

tα	Critical value
t$_0$	Test statistic
µ	The population mean, which is generally not known
UCL$\bar{x}$	Upper confidence limit of mean
$\bar{x}$	The sample mean, which is determined from the measurements taken
x$_i$	The ith measurement in the dataset

8.3. Glossary

α risk

The probability, expressed as a decimal, of making a Type I error when a hypothesis is tested statistically. A Type I error wrongly rejects a null hypothesis when in fact the null hypothesis is true. In this document, the null hypothesis always assumes that the site is contaminated and thus the α risk refers to the probability of a site being validated as uncontaminated when it is contaminated.

β risk

The probability, expressed as a decimal, of making a Type II error when a hypothesis is tested statistically. A Type II error wrongly accepts a null hypothesis when in fact the null hypothesis is false. In this document, the null hypothesis always assumes that the site is contaminated and thus the β risk refers to the probability that a site is determined as contaminated when it is uncontaminated.

Acceptable limit

A threshold concentration value below which the level of contamination is regarded as acceptable. An acceptable limit can either be adopted from the appropriate guidelines or it can be derived on a site-specific basis using risk assessment. Where site remediation is involved, acceptable limits are often referred to as 'clean-up standards' or 'remediation standards'.

Acceptance criteria

A statistical statement specifying how a contaminant distribution will be compared with an acceptable limit (see above definition) to determine whether a site should be evaluated as 'contaminated' or 'uncontaminated'. Contaminant concentrations can vary over orders of magnitude in a sampling area. All site assessments must state the appropriate acceptance criteria, as well as the appropriate acceptable limits.

Ambient air

External air environment, not including the air inside buildings or structures.

Arithmetic mean

The arithmetic mean is commonly referred to as the average and is used to describe the centre of the data distribution. It is obtained by adding all the values and dividing the result by the number of values.

Central tendency

The central or typical value for a probability distribution, and may be considered the average value in a dataset. It is generally described as the mode, median, or, more commonly, the mean, and describes where a sample distribution is centred.

Chi-squared distribution

A type of cumulative probability distribution that varies depending on the degrees of freedom (df). It is used to test relationships between categorical variables in the same population.

Coefficient of variation (CV)

CV is the measurement of the relative homogeneity of a distribution. Low CV values, for example 0.5 or less, indicate fairly homogeneous contaminant distribution, while CVs with values of more than 1–1.2 imply that the concentration distribution of a contaminant is heterogeneous and probably highly skewed to the right.

Composite sample

The bulking and thorough mixing of soil samples collected from more than one sampling location to form a single soil sample for chemical analyses.

Conceptual site model (CSM)

Provides a three-dimensional overview of the contamination at sites and their surroundings, highlighting the sources, receptors and exposure pathways between the sources and receptors.

Confidence level

The probability, expressed as a percentage, that a statistical statement is correct. Confidence level is the opposite expression of 'risk' (see definitions of α and β risks). In this document when a risk that needs to be regulated, the confidence level is always equal to $I - \alpha$.

Contaminated

For the purpose of this document and depending on the context, 'contaminated' can have slightly different meanings. If a site or a sampling area is evaluated as contaminated, it means that it has not met the acceptance criteria (see definition of acceptance criteria). Contaminated can also describe a localised area or soil that has contaminant concentrations exceeding an acceptable limit (see definition of acceptable limit). Note: depending on what the acceptance criteria are, an entire site could be considered uncontaminated even though a certain percentage of it is expected to be contaminated.

Contamination

NEPC 2013 defines contamination as 'the condition of land or water where any chemical substance or waste has been added as a direct or indirect result of human activity at above background level and represents, or potentially represents, an adverse health or environmental impact'.

Data quality objectives (DQOs)

A systematic planning process that defines the type, quantity and quality of data needed to support decisions on the environmental condition of a site or a specific decision area.

Decision area

A specific area or medium on a site, or offsite, about which data is being gathered so a decision can be made. For example, a decision can be made on part of a site, soil, a stockpile, soil gas, groundwater, surface waters or sediments.

Estimate

An estimate is a value that is inferred for a population based on data collected from a sample of units from that population. For example, the measured data from a sampling event used to calculate the sample mean ($\bar{x}$) is then used to estimate the population mean (μ).

Estimation

A technique that systematically adjusts the sample data to determine an estimated value for the population.

Geometric mean

This is similar to the **arithmetic mean** (described above), in that it is also a measure of the central tendency of the distribution of a population or sample. It is sensible to calculate geometric means only on

populations or samples that contain positive values. The geometric mean is obtained by multiplying n values from the dataset together, then taking the nth root of the product.

Grab samples

Samples collected from different locations that will be analysed individually.

Hotspot

A localised area where the level of contamination is noticeably greater than in surrounding areas. Note that a hotspot is only **relatively** high in contamination.

Inter well

Comparison between two groundwater monitoring wells that are separated spatially.

Intra well

Comparison of measurements over time at one groundwater monitoring well.

Maximum

The maximum observed value in a data. It generally provides a conservative estimate of the potential exposure risks so if the maximum is below the action level, the site should be suitable for its proposed land use.

Median

The middle value of the distribution. Half the data values are less than the median and half are greater.

Minimum size effect

The acceptable magnitude of the difference between the populations or groups being studied.

Mode

The value that occurs most frequently. It is determined by counting the number of times each value occurs.

Modules

A series of discrete DQOs outputs, based on logical categories, that addresses selected components of a site investigation. Modules can be selected for contaminant types, media, decision areas, or a workable combination of these.

Neyman–Pearson method

A method of statistical inference used to determine if a null hypothesis (H_0) should be rejected in favour of an alternative hypothesis (H_A), at a specified level of confidence.

Outlier

A data point that sits outside the expected range of the data. An outlier can have either a high or low value. Outliers must be retained in sample datasets unless there is a demonstrable reason for rejecting them such as a coding error, sample contamination or equipment failure.

Parameters

Numerical measures of the characteristic of interest in the population being sampled. Typical parameters are the population mean (μ), variance (σ^2) and standard deviation (σ). Parameter values are usually unknown.

Percentiles and quartiles

These are descriptive values used to equally split a dataset into 100 parts. A percentile is the value that a percentage of observations in a dataset is equal to or less than, for example, 80% of observations in a dataset are at or below the 80th percentile, while 20% are above.

Quartiles are commonly used to break the dataset up into four equal parts, providing an indication of the distribution and variance of the data.

- First quartile – the 0th percentile up to (and including) the 25th percentile.
- Second quartile – from the 25th percentile up to (and including) the 50th percentile.
- Third quartile – from the 50th percentile up to (and including) the 75th percentile.
- Fourth quartile – from the 75th percentile up to (and including) the 100th percentile.

Population

Any large collection of objects, things or individuals with some characteristics in common, that is being studied and for which information is sought. The population under consideration must be clearly and succinctly defined to allow effective sampling design and subsequent reporting.

The population can be further defined as the **target population** and the **sampled population**, and ideally these should be the same. The target population is the set of all units that comprise the items of interest, that is the population about which a decision is required, and the sampled population is that part of the target population that is accessible and available for sampling. If the two diverge significantly, the target population should be redefined.

Probabilistic sampling

Probabilistic sampling occurs when each member of the population has a given probability (greater than zero and less than one) of being included in the sample. If the probability is the same for all population members the sample will be unbiased. Because inclusion in the sample is based on probability, subsequent samples will not necessarily include the same members.

Range

The range of a dataset measures the spread between the highest and lowest values in it. Other measures such as the standard deviation and the interquartile range are required to provide an understanding of the data's distribution.

Residual soil

The soil at a site that is not contaminated by industrial, commercial, or agricultural activities, consistent with the term 'ambient background concentration' (ABC) from the NEPM. Residual soils can include natural soils, reworked natural soils and historically imported material. Residual soils may have naturally occurring background levels of contaminants, contaminants that have been introduced from diffuse or non-point sources by general anthropogenic activity, and only low levels of contaminants attributed to industrial, commercial, or agricultural activities.

Sample

'Sample' has several meanings including:

- as more broadly used in statistics, a representative group drawn from a population for description or measurement
- a physical amount of a material such as soil, water or air or an aliquot, taken for testing or chemical analysis
- a sampling point or sample location, being the location in plan at which a sample is collected, including description, for example, geological logs and field screening, for example, PID or XRF.

Sample size

The number of samples or sampling points in a sampling program.

Sampling, analysis and quality plan (SAQP)

Incorporates the CSM and the DQO outputs, to provide the context of and justification for the selected sampling and analysis. The methods, procedures and QC samples associated with the DQIs, including the frequency and MQOs and any associated contingencies, are also documented. The SAQP ensures that the data collected is representative and provides a robust basis for site assessment (NEPC 2013).

Sampling pattern

The locational pattern of sampling points in a sampling area.

Sampling point

The location at which a sample is collected.

Site characterisation

The assessment of the nature, level and extent of contamination. A typical site characterisation involves a preliminary site investigation (PSI), followed by a detailed site investigation (DSI), where warranted.

Site validation

The process of showing that a site is successfully remediated.

Standard deviation

Calculated by taking the square root of the variance (described below). It provides an indication of a population or sample data's typical deviation from its mean.

Statistic

Any summary number that describes the sample, such as an average or percentage. For example, the mean of a sample is described as $\bar{x}$ (x-bar) and the standard deviation as **s**. When describing the population from which the sample is drawn, a summary number is called a **parameter**.

Statistical power

The probability of correctly determining a positive result based on sample data, for example, a change or difference in the population.

Sub-sample

A sample that will be combined with other sub-samples to form a composite for chemical analyses.

Systematic planning

A planning process based on a scientific method which helps the project to unfold logically. Systematic planning includes established management and scientific elements. In the assessment of site contamination, it includes the application of the **DQOs** process and development of a **CSM** and **SAQP**.

Variable

A characteristic, number or quantity that is the subject of the inquiry. In the assessment of site contamination, it is usually continuous numerical variables that are being assessed, for example the concentration of a contaminant in soil, soil gas or water. Discrete or discontinuous variables are at times considered, such as the number of fish in a waterbody. These are both quantitative variables in that they are derived by measurements.

Qualitative or categorical variables include ordinal or ranked variables and nominal variables. Ordinal variables are observations that take a value that can logically be ordered or ranked, such as first, second, third, whereas nominal observations take a value that cannot be organised in a logical sequence, such as presence or absence. Categorical variables are not commonly used in the assessment of site contamination.

Variance

The average squared distance of population or sample data points from the associated mean.

Weight of evidence/lines of evidence

'Weight of evidence' describes the process of collecting, analysing and evaluating a combination of different qualitative, semi-quantitative or quantitative lines of evidence to make an overall assessment of contamination.

Applying a weight-of-evidence process incorporates judgements about the quality, quantity, relevance and congruence of the data contained in the different lines of evidence (ANZG 2018).

9. References

Australian and New Zealand Environment and Conservation Council (ANZECC) and Agriculture and Resource Management Council of Australia and New Zealand (ARMCANZ) 2000, *Australian and New Zealand guidelines for fresh and marine water quality*, paper no. 4, ANZECC and ARMCANZ, Canberra.

Australian and New Zealand Governments and Australian State and Territory Governments (ANZG) 2018, *Australian and New Zealand guidelines for fresh and marine water quality*, Water Quality Australia, Canberra ACT, www.waterquality.gov.au/anz-guidelines.

British Standards Institution (BSI) 2013, *Investigation of potentially contaminated sites: code of practice*, BS 10175:2011+A1:2013, BSI Standards Limited.

Clements L, Palaia T & Davis J 2009, *Characterisation of sites impacted by petroleum hydrocarbons: national guideline document*, CRC Care Technical Report no. 11, CRC for Contamination Assessment and Remediation of the Environment, Adelaide.

Contaminated Land: Applications in Real Environments (CL:AIRE) 2008, *Guidance on Comparing Soil Contamination Data with a Critical Concentration*, The Chartered Institute of Environmental Health and CL:AIRE, London.

CRC Care 2013, *Petroleum hydrocarbon vapour intrusion assessment: Australian guidance*, CRC CARE Technical Report no. 23, CRC for Contamination Assessment and Remediation of the Environment, Adelaide.

Crumbling DM 2002, In search of representativeness: evolving the environmental data quality model, *Quality Assurance*, vol.9, pp179–90, http://cluin.org/download/char/dataquality/dcrumbling.pdf.

Davis GB, Wright J & Patterson BM 2009, *Field Assessment of vapours,* CRC CARE Technical Report no. 13, CRC for Contamination Assessment and Remediation of the Environment, Adelaide.

Department of Agriculture and Water Resources 2018, *Australian & New Zealand guidelines for fresh and marine water quality*, Department of Agriculture and Water Resources, Canberra, www.waterquality.gov.au/anz-guidelines.

Department of Environment and Conservation (DEC) 2004, New South Wales (NSW) *Australian river assessment system (AUSRIVAS) sampling and processing manual*, DEC NSW, Sydney.

Department of Environment and Conservation (DEC) 2005a, *Contaminated sites: Guidelines for assessing former orchards and market gardens*, DEC 2005/195, DECCW NSW, Sydney.

Department of Environment and Conservation (DEC) 2005b, *Information for the assessment of former gasworks sites*, DEC 2005/237, DECCW NSW, Sydney.

Department of Environment and Conservation (DEC) 2007, *Contaminated sites: Guidelines for the assessment and management of groundwater contamination*, DEC 2007/144, DEC NSW, Sydney.

Department of Environment, Climate Change and Water (DECCW) 2009, *Guidelines for implementing the Protection of the Environment Operations (Underground Petroleum Storage Systems) Regulation 2008*, DECCW 2009/653, DECCW NSW, Sydney.

Department of Environment, Climate Change and Water (DECCW) 2010, *Vapour intrusion: technical practice note*, DECCW 2010/774, DECCW NSW, Sydney.

Department of Environment (DoE) Queensland 1998, *Draft guideline for the assessment & management of contaminated land in Queensland*, DoE, Brisbane.

Department of Environment and Science (DES) Queensland 2018, *Monitoring and sampling manual: environmental protection (water) policy 2009* [sic], DES, Brisbane.

Department of Health and Ageing and EnHealth Council 2012, *Environmental health risk assessment: guidelines for assessing human health risks from environmental hazards*, Department of Health and Ageing, Canberra.

Devore J & Farnum N 2005, *Applied statistics for engineers and scientists*, 2nd Edition, Brooks/Cole, Cengage Learning, Belmont CA.

Environment Protection Authority (EPA) 1995, *Contaminated sites: Guidelines for the vertical mixing of soil on former broad-acre agricultural land*, EPA 2003/28, NSW EPA, Sydney.

Environment Protection Authority (EPA) 1997, *Contaminated sites: guidelines for assessing banana plantation sites*, EPA 97/37, NSW EPA, Sydney.

Environment Protection Authority (EPA) 2014a, *Resource Recovery Order under Part 9, Clause 93 of the Protection of the Environment Operations (Waste) Regulation 2014: the excavated natural material order 2014.*

Environment Protection Authority (EPA) 2014b, *Best practice note: landfarming*, EPA 2014/0323, NSW EPA, Sydney.

Environment Protection Authority (EPA) 2015a, *Guidelines on the duty to report contamination under the Contaminated Land Management Act 1997*, EPA 2015/0164, NSW EPA, Sydney.

Environment Protection Authority (EPA) 2015b, *Technical note: light non-aqueous phase liquid assessment and remediation*, EPA 2015/0553, NSW EPA, Sydney.

Environment Protection Authority (EPA) 2017, *Contaminated land management: guidelines for the NSW site auditor scheme (3rd edition)*, EPA 2017P0269, NSW EPA, Sydney.

Environment Protection Authority (EPA) 2018, *Guidelines on resource recovery orders and exemptions: for the land application of waste materials as fill*, EPA 2017/P0392, NSW EPA, Sydney.

Environment Protection Authority (EPA) 2020a, *Assessment and management of hazardous ground gases: contaminated land guidelines*, EPA 2019P2047, NSW EPA, Sydney.

Environment Protection Authority (EPA) 2020b, *Consultants reporting on contaminated land: contaminated land guidelines*, EPA 2020P2233, NSW EPA, Parramatta.

Environment Protection Authority (EPA) 2020c, *Guidelines for implementing the Protection of the Environment Operations (Underground Petroleum Storage Systems) Regulation 2019*, EPA 2020/P2700, NSW EPA, Parramatta.

Environment Protection Authority (EPA) South Australia 2005, *Composite soil sampling in site contamination assessment and management*, Government of South Australia, Adelaide.

Environment Protection Authority (EPA) Victoria 2009, *Industrial waste resource guidelines: soil sampling*, IWRG702.

Ferguson CC 1992, The statistical basis for spatial sampling of contaminated land*, Ground Engineering*, vol. 25, no. 6, pp 34–38.

Gilbert RO 1987, *Statistical methods for environmental pollution monitoring*, John Wiley & Sons Inc., Brisbane.

Gray JM & Murphy BW 1999, *Parent material and soils: A guide to the influence of parent material on soil distribution in eastern Australia*, Technical Report No. 45, NSW Department of Land and Water Conservation, Sydney.

Hamon RE, Mclaughlin MJ, Gilkes RJ, Rate AW, Zarcinas B, Robertson A, Cozens G, Radford N & Bettenay L 2004, Geochemical indices allow estimation of heavy metal background concentrations in soils, *Global Biogeochemical Cycles*, vol. 18, GB1014.

Harr ME 1987, *Reliability-based design in engineering*, McGraw-Hill, New York.

HEPA 2018, *PFAS national environmental management plan version 2.0*, Heads of EPAs Australia and New Zealand.

Interstate Technology and Regulatory Council (ITRC) 2007, *Vapor intrusion pathway: a practical guideline*, VI-1, ITRC Vapor Intrusion Team, Washington DC, USA, www.itrcweb.org/Documents/VI-1.pdf.

Interstate Technology and Regulatory Council (ITRC) 2012, *Incremental sampling methodology*, (ISM-1), ITRC, Washington DC, USA, https://itrcweb.org/teams/projects/incremental-sampling-methodology.

Interstate Technology and Regulatory Council (ITRC) 2013, *Groundwater statistics and monitoring compliance: statistical tools for the project life cycle*, GSMC-1, ITRC, Washington DC, USA, http://www.itrcweb.org/gsmc-1/.

Interstate Technology and Regulatory Council (ITRC) 2017, *Naming conventions and physical and chemical properties per- and polyfluoroalkyl substances (PFAS)*, ITRC, Washington DC.

Lock WH 1996, Composite sampling, in *National Environmental Health Forum Monographs, Soil Series No. 3*, South Australian Health Commission, Adelaide.

McDougall KW & Macoun TW 1996, *Guidelines for the assessment and clean up of cattle tick dip sites for residential purposes*, NSW Agricultural in conjunction with CMPS&F Environmental, Wollongbar NSW.

Naidu R, Jit J, Kennedy B & Arias V 2016, Emerging contaminant uncertainties and policy: The chicken or the egg conundrum, *Chemosphere*, vol. 154, pp 385–390.

National Environment Protection Council (NEPC) 2013, *National environment protection (assessment of site contamination) amendment measure 2013 (no. 1)*, Schedule A and Schedules B(1)–B(9), National Environment Protection Council, Canberra.

Nickerson RS 2000, Null hypothesis significance testing: A review of an old and continuing controversy, *Psychological Methods*, vol. 5, no. 2, pp 241–301.

New Jersey Department of Environmental Protection (NJDEP) 2005, *Vapor intrusion guidance*.

Northern Territory Environment Protection Authority (EPA) 2013, *Guidelines on conceptual site models*, NT EPA, Darwin.

Office of Environment and Heritage 2011, *Guidelines for consultants reporting on contaminated sites*, OEH 2011/0650, NSW OEH, Sydney.

Parkhurst DF 1998, Arithmetic versus geometric means for environmental concentration data, *Environmental Science & Technology*, vol. 32 (3), pp 92A–98A.

Perezgonzalez JD 2015, Fisher, Neyman–Pearson or NHST? A tutorial for teaching data testing, *Frontiers in Psychology, vol. 6*, article 223, p. 1.

Provost LP 1984, Statistical methods in environmental sampling, in Schweitzer GE and Santolucito JA (eds), *Environmental sampling for hazardous wastes*, American Chemical Society, Washington DC.

Reinhart A 2015, *Statistics done wrong: The woefully complete guide*, No Starch Press, San Francisco CA.

Simpson S and Batley G (eds) 2016, *Sediment quality assessment: a practical guide*, CSIRO Publishing, Melbourne.

South Australian Health Commission (SAHC) 1995, *Guidelines for the composite sampling of soils*, SAHC, Adelaide.

US Environmental Protection Agency (USEPA) 1996, *Soil screening guidance: user's guide* (2nd edition), Attachment B, Soil Screening DQOs for Surface Soils and Subsurface Soils, EPA/540/R-96/018, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 2000, *Data quality objectives process for hazardous waste site investigations (QA/G-4HW)*, EPA/600/R-00/007, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 2001, *Guidance on data quality indicators (QA/G-5i)*, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 2002a, *Guidance on environmental data verification and data validation (QA/G-8)*, EPA/240/R-02/004, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 2002b, *Guidance on choosing a sampling design for environmental data collection for use in developing a quality assurance project plan (QA/G-5S)*, EPA/240/R-02/005, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 2002c, *Guidance for quality assurance project plans (QA/G-5)*, EPA/240/R-02/009, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 2002d, *Calculating upper confidence limits for exposure point concentrations at hazardous waste sites*, OSWER 9285.6-10, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 2006a, *Data quality assessment: statistical methods for practitioners (QA/G-9S)*, EPA/240/B-06/003, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 2006b, *Guidance on systematic planning using the data quality objectives process (QA/G-4)*, EPA/240/B-06/001, Appendix: Derivation of sample size formula for testing mean of normal distribution versus an action level, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 2006c, *Data quality assessment: a reviewer's guide (QA/G-9R)*, EPA/240/B-06/002, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 2007, *Guidance for preparing standard operating procedures (SOPs) (QA/G-6),* EPA/600/B-07/001, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 2009, *Statistical analysis of groundwater monitoring data at RCRA facilities: Unified Guidance*, EPA 530/R-09-007, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 2014, *Test methods for evaluating solid waste: physical/chemical methods compendium (SW-846)*, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 2015a, *ProUCL version 5.1.002: technical guide: statistical software for environmental applications for datasets with and without nondetect observations*, EPA/600/R-07/041, USEPA, Washington DC.

US Environmental Protection Agency (USEPA) 2015b, *ProUCL version 5.1.002: User guide: Statistical Software for environmental applications for datasets with and without nondetect observations, EPA/600/R-07/041*, USEPA, Washington DC.

Viveros R 1997, Inference about the mean in log-regression with environmental applications, *Environmetrics*, vol. 8(5), pp. 569–582.

Wendelberger J & Campbell K 1994, *Non-detect data in environmental investigations*, American Statistical Association, Toronto, Canada.

Wilson S, Card G & Haines S 2009, *Ground gas handbook*, Whittles Publishing, Dunbeath, UK.

Appendix A: Descriptive statistics

This Appendix briefly reviews the descriptive statistics commonly used for summarising data.

Appendices B to D show how they are used, giving specific procedures and worked examples.

Range and percentiles

The range of a dataset measures the spread between the highest and lowest observed values in the dataset. It can be expressed as an interval, such as a–b, where a is the lowest value and b is the highest, or it can be expressed as an interval width, such as b - a = c. While either approach covers the range of the observed values, the maximum value is of particular concern in the assessment of site contamination, and the range is generally more informative as an interval, as it shows the spread and the extremes of the data.

As the range only measures the spread between highest and lowest values, other measures such as the **standard deviation** or the **interquartile range** (IQR) are needed to more fully describe data distribution.

Maximum

The maximum observed value in a dataset is important when assessing site contamination as it generally conservatively estimates the potential exposure risks. It is usually assumed that if the maximum is below the action level, the site will be suitable for the associated land use. However, this assumption is true only if there is enough data and the data is representative. If this is not the case, the maximum observed value may overestimate or underestimate the risk.

Where the consequences of decision error will be severe, or there are not enough samples to estimate the population mean from the sample mean, the maximum value can be used as an estimate of the population mean and termed **max test** for statistical analysis. This is often done for judgmental samples, such as with soil gas or groundwater data. Where this approach is used, it should be appropriately documented and justified.

Percentiles and quartiles

Percentiles, as suggested by the name, are descriptive values used to equally split a dataset into 100 parts. The Xth percentile in a dataset has a value greater than or equal to X% of the data – for example, the 80th percentile has a value greater than or equal to 80% of the data.

Percentiles can be used as the statistical parameter of interest, for instance, for comparing to criteria or action levels. For example, ANZG 2018 states that '[f]or toxicants, it is recommended that action is triggered if the 95th percentile of the test data exceeds the guideline value'.

Quartiles are used to break up the dataset into four equal parts, providing an indication of the distribution and variance of the data. When observations are placed in ascending order by value:

- the first quartile, Q1, also called the lower quartile, is the value of the observation at or below which a quarter (25%) of observations lie, and is the 25th percentile
- the second quartile, Q2, is the median value at or below which half (50%) of observations lie, and is the 50th percentile
- the third quartile, Q3, also called the upper quartile, is the value of the observation at or below which three-quarters (75%) of the observations lie and is the 75th percentile.

The interquartile range (IQR) is used as a measure of the spread of the dataset which also indicates its dispersion. It is the difference between the upper and lower quartiles (Q3 - Q1 = IQR), that is, it measures the spread between the 25th and 75th percentiles. The IQR spans 50% of a dataset and eliminates the influence of outliers as it excludes the highest and lowest quarters.

Percentiles and quartiles can be used for datasets with limited observations, and for all types of data collection, as their use requires no assumptions about the underlying distribution or whether the samples were judgmental or probabilistic. However, ANZG 2018 notes that the precision with which percentiles are estimated depends heavily on the sample size, with at least 13 samples need to estimate the 25th

and 75th percentiles with an associated 95% confidence interval, and a minimum of 36 samples needed to estimate the 10th and 90th percentiles. Even larger sample sizes are required to estimate extreme percentiles, i.e. the 5th and 95th.

As the IQR does not depend on extreme values, it can be used when a dataset includes non-detects, at least where < 25% of the data is below the limits of reporting (LORs). For datasets that are not nearly-normal, or which contain extreme values, the IQR may be more representative of the dispersion of the data than the standard deviation. The IQR is therefore described as a robust estimate.

Appendix B explains how to determine quartiles.

Central tendency

Central tendency is the central or typical value for a probability distribution, and may be considered the average value in a dataset. It is generally described by the mode, median, or, most commonly, the mean, and indicates where a sample distribution is centred. While these estimates can generally be regarded as being representative or typical of the data, for small or highly skewed datasets they should be considered as approximations only.

Appendix C explains how to determine measures of central tendency.

Mode

The mode is the value that occurs most frequently and is determined by counting the number of times each value occurs. Since a sample mode may not exist or may not be unique, for example, the distribution may be bimodal, it is rarely used as a measure of central tendency, although it can be useful for qualitative data such as categories.

Median

The median is the middle value of the distribution: half the data points have values greater than the median, and half have values less than it. The sample median is not influenced by extreme values so can be used when the underlying distribution is unknown: it is commonly used to describe the centre of the distribution when non-parametric methods are employed. The median can also be used if non-detects are present, although care should be taken if there are many of them. If a median is found to be a non-detect while there are locations reporting values above detection levels, stratifying the site should be considered.

A number of guidelines recommend the use of median values in certain circumstances. For example:

- NEPC 2013, B2 states that when using non-parametric approaches, the median can be used to describe the centre of the distribution
- ANZG 2018 notes that for comparing test data with guideline values for physico-chemical stressors, '[a] trigger for further investigation of the test water body will be deemed to have occurred when the median concentration of a particular measurement parameter in **n** independent samples taken at the test water body exceeds the 80th percentile (or is below the 20th percentile if "less is worse") of the same measurement parameter at the reference site'.

Arithmetic mean

The arithmetic mean is commonly referred to as the average and is used to describe the centre of the data distribution. The arithmetic mean is denoted as **μ** (lowercase Greek letter *mu*) for the population mean or as $\bar{x}$ (x-bar) for the sample mean. In the assessment of site contamination, the population mean is generally not known, so the sample mean is used as an estimate of the population mean. The arithmetic mean is calculated by dividing the sum of the sample measurements by the number of samples.

Larger sample sizes tend to produce sample means that are closer to the population mean, as in theory extreme data values balance each other out. But when sample sizes are small, the arithmetic mean can be affected by outliers, and when judgmental sampling is used, the arithmetic mean is often a biased measure of central tendency.

The mean value may be more representative of site contamination than the maximum value by providing a better estimate of the contaminant concentrations that receptors would be exposed to over a period of time. However, it is important that small areas of high concentration (hotspots) are not ignored by averaging with lower values from other parts of the site or the decision area.

Geometric mean

The geometric mean is similar to the arithmetic mean in that it is also a measure of the central tendency of the distribution of a population or sample. This is also described as the arithmetic mean of the logarithmic scale of a dataset, or the nth root of the product of n numbers.

Due to the log transformation involved in the calculation, the geometric mean is not as affected by outliers as the arithmetic mean, and is commonly used when the data is skewed or log-normally distributed. However, the curvature of the logarithmic function may downplay the higher values in favour of the lower ones.

Higher values are important in the assessment of site contamination. If assumptions regarding the condition of a site are based on the geometric mean, downplaying higher values may increase the chance of a Type I error. Because of this potential bias, geometric means, including back transformation, should not be used in isolation to compare against action levels, and if they are used appropriate justification should be provided. Where log-transformed data are approximately normal or at least reasonably symmetric, back transformation may be appropriate (USEPA 2009 and Viveros 1997), but for skewed datasets that are not log-normal, the geometric mean is likely to be a poor estimator of population mean (Parkhurst 1998).

Variability

An important aspect of data analysis is determining the variability of the sampling data. Calculating variability can provide an indication of how heterogenous the variables are likely to be across a decision area, and how representative of the sampling data the measures of central tendency are. The variability of data is measured by **variance**, **standard deviation** and the **coefficient of variation**.

See Appendix D for how to determine measures of variability.

Variance

Variance is the average squared distance of each data point from the sample mean. It can be affected by extreme values and by large numbers of values below the LORs.

Standard deviation

The standard deviation is calculated by taking the square root of the variance and provides an indication of the data's typical deviation from the mean. The standard deviation of a population is denoted as **σ** (Greek lowercase sigma), and as **s** for a sample. The sample standard deviation is commonly used in site contamination assessment, as the standard deviation of the population is generally not known.

A large sample variance or standard deviation indicates that the data points are not closely clustered around the mean. Both the variance and the standard deviation are strongly influenced by the number of samples collected, and influenced by extreme values in either direction.

Coefficient of variation

The coefficient of variation (CV) or relative standard deviation (RSD) measures the relative homogeneity of a distribution. The CV is determined as the standard deviation of a distribution divided by the mean of the distribution, that is, $CV = s/\bar{x}$ for sample data. The RSD is determined in the same way, but expressed as a percentage, that is, a CV of 0.5 = an RSD of 50%.

Low CV values, for example, of 0.5 or less, indicate a fairly homogeneous contaminant distribution, while CVs with values of more than 1–1.2 imply that the concentration distribution of a contaminant is heterogeneous.

Appendix B: Determining quartiles

Percentiles are descriptive values that split a dataset into 100 equal parts, providing a representation of the sampling data that can be used for either normal or non-normal distributions. A percentile provides the value that a given percentage of observations in a dataset is less than or equal to (for example, 25% of observations in the dataset have values at or below the value of the 25th percentile).

Percentiles can be used in the statistical analysis of datasets that have limited observations. The dataset can also be divided by **quartiles**, which are the 25th, 50th and 75th percentiles.

Determination

To calculate percentiles, values are ordered from the lowest to the highest and assigned a rank, with the required percentile calculated using the formula shown below. While this procedure can be used for small datasets, it is commonly conducted using spreadsheets or statistical packages. Note that all percentiles of sample data are biased estimators of population percentiles.

The values are ranked from lowest to highest:

$$X_{(1)}, X_{(2)}, X_{(3)}, X_{(4)} \ldots, X_{(n)}$$

The pth percentile is calculated by:

$$y_p = (1 - f) \times X_i + f \times X_{(i+1)}$$

Where:

y_p	the value of the pth percentile
p^{th}	the specified percentile
r	$(n - 1)p + 1$
floor(r)	calculate r and discard decimals
i	floor(r)
f	r – i
X_i	the value of the ith rank
$X_{(i + 1)}$	the value of the ith + 1 rank

The data in Table 1 is used for the worked examples in this and the following appendices.

Table 1   Summary of analytical results – metals in soil (mg/kg)

Sample ID or statistic	Arsenic	Chromium	Copper	Lead	Nickel	Zinc
Limits of reporting	5	2	5	5	2	5
Analytical						
Analytical sample B2-01	103	12	34	20	18	11
Analytical sample B2-02	50	21	30	7	2	10
Analytical sample D2-01	43	26	83	17	14	35
Analytical sample D2-02	9	10	29	14	5	12
Analytical sample A4-01	203	4	260	18	12	232
Analytical sample A4-02	54	5	55	17	9	41
Analytical sample C4-01	341	19	401	133	7	543
Analytical sample C4-02	34	17	46	16	10	13
Analytical sample B6-01	71	18	24	14	5	9
Analytical sample B6-02	14	6	8	17	12	5
Analytical sample D6-01	62	11	51	15	3	36
Analytical sample D6-02	6	4	18	16	24	10
Analytical sample A8-01	27	17	61	16	4	24
Analytical sample A8-02	7	10	38	20	13	10
Analytical sample C8-01	24	15	39	12	6	8
Analytical sample C8-02	13	16	17	14	19	7
Descriptive statistics						
Number of samples	16	16	16	16	16	16
Number of detects	16	16	16	16	16	16
Percentage non detects	0%	0%	0%	0%	0%	0%
Maximum	341	26	401	133	24	543
Third quartile	64.3	17.3	56.5	17.3	13.3	35.3
Median value	38.5	13.5	38.5	16.0	9.4	11.5
First quartile	13.8	9.0	27.8	14.0	5.2	9.8
Minimum	6	4	8	7	2	5
Arithmetic average	66.3	13.2	74.6	22.9	10.2	62.9
Geometric average	35.2	11.4	43.5	17.3	8.3	20.0
Mode	-	10	-	17	12	10
Variance	7,792.2	42.4	10,988.8	872.1	39.7	19,410.1
Standard deviation	88.3	6.5	104.8	29.5	6.3	139.3
Coefficient of variation (CV)	1.3	0.5	1.4	1.3	0.6	2.2
Inferential statistics						
Standard error of the mean (SE$\bar{x}$)	22.1	1.6	26.2	7.4	1.6	34.8
Relative standard deviation (RSD)	133.1%	49.4%	140.5%	129.1%	61.9%	221.6%
Margin of error (MoE)	47.0	3.5	55.9	15.7	3.4	74.2

Sample ID or statistic	Arsenic	Chromium	Copper	Lead	Nickel	Zinc
Maximum probability error (MPE)	70.9%	26.3%	74.9%	68.8%	33.0%	118.1%
95% UCLx̄ two-sided Student's t	113.4	16.7	130.5	38.6	13.5	137.1
95% UCLx̄ one-sided Student's t	105.0	16.0	120.5	35.8	13.0	123.9
ProUCL determination	120.5	16.0	135.2	55.1	13.0	214.7
Method recommended*	Gamma	Student's t	H-UCL	Chebyshev	Student's t	Chebyshev
Criteria and number of samples						
HIL-A land use (NEPC 2013, B1)	100	100	6,000	300	400	7,400
Number of samples to be used (whole number) – CRV method	44	2	2	2	2	2
Number of samples – MPE method	15	18	16	16	14	15

Worked example

The metals data in mg/kg from Table 1 is used in this example.

To determine the 25th percentile of the sampling data for arsenic (As):

The values are ordered from lowest to highest and assigned a rank:

$$X_{(1)} = 6, X_{(2)} = 7, X_{(3)} = 9, \mathbf{X_{(4)} = 13}, \mathbf{X_{(5)} = 14}, X_{(6)} = 24, X_{(7)} = 27, X_{(8)} = 34,$$

$$X_{(9)} = 43, X_{(10)} = 50, X_{(11)} = 54, X_{(12)} = 62, X_{(13)} = 71, X_{(14)} = 103, X_{(15)} = 203, X_{(16)} = 341$$

Bolded values are $X_{(i)}$ and $X_{(i+1)}$.

The input parameters are calculated for the 25th percentile:

$$r = (n - 1)p + 1$$

$$r = (16 - 1)\,0.25 + 1$$

$$r = 4.75$$

$$i = 4$$

$$f = r - i$$

$$f = 0.75$$

The 25th percentile is calculated as:

$$y_{(0.25)} = (1 - f) \times X_i + f \times X_{(i+1)}$$

$$y_{(0.25)} = (1 - 0.75) \times 13 + 0.75 \times 14$$

$$y_{(0.25)} = 13.8$$

The 25th percentile of the sampling data for As is 13.8 mg/kg.

Appendix C: Determining measures of central tendency

The **central tendency** is a central or typical value for a probability distribution, and may be considered the average value in a set of data. Methods for calculating the **median**, the **arithmetic mean** and the **geometric mean** are shown below.

The **mode** is the value that occurs with the greatest frequency (that is, the greatest number of times): to calculate it, simply count the number of times each value occurs. As the mode does not always exist or may not be unique, it is the value of central tendency that is least commonly used, although it can be useful for describing qualitative data.

Determination

Measures of central tendency are determined as follows.

Median with an odd number of samples

$$median = X_{(n+1)/2}$$

Median with an even number of samples

$$median = \frac{1}{2}\left[X_{(n/2)} + X_{(n/2+1)}\right]$$

Arithmetic mean

$$sample\ arithmetic\ mean\ = \frac{(X_1 + X_2 + \ldots X_n)}{n}$$

Geometric mean

$$sample\ geometric\ mean = \sqrt[n]{(X_1 \times X_2 \times \ldots X_n)}$$

Worked example

The metals data in mg/kg from Table 1 is used in this example to determine the measures of central tendency for the sampling data for arsenic (As):

Median

The values are ordered from lowest to highest and assigned a rank:

$X_{(1)} = 6$, $X_{(2)} = 7$, $X_{(3)} = 9$, $X_{(4)} = 13$, $X_{(5)} = 14$, $X_{(6)} = 24$, $X_{(7)} = 27$, **$X_{(8)} = 34$, $X_{(9)} = 43$,**

$X_{(10)} = 50$, $X_{(11)} = 54$, $X_{(12)} = 62$, $X_{(13)} = 71$, $X_{(14)} = 103$, $X_{(15)} = 203$, $X_{(16)} = 341$

Bolded values are $X_{(n/2)}$ and $X_{(n/2 + 1)}$.

As n = 16, an even number, the sample median is determined as:

$$\text{sample median} = \frac{1}{2}\left[X_{\left(n/2\right)} + X_{\left(n/2 + 1\right)}\right]$$

$$\text{sample median} = \frac{1}{2}\left[X_{\left(16/2\right)} + X_{\left(16/2 + 1\right)}\right]$$

$$\text{sample median} = \frac{1}{2}\left[X_{(8)} + X_{(9)}\right]$$

$$\text{sample median} = \frac{1}{2}\left[34 + 43\right]$$

$$\text{sample median} = 38.5$$

The sample median for As is 38.5 mg/kg.

Arithmetic mean

$$\text{sample arithmetic mean} = \frac{(X_1 + X_2 + \dots X_n)}{n}$$

$$\text{sample arithmetic mean} = \frac{(103 + 50 + \dots 13)}{16}$$

$$\text{sample arithmetic mean} = 66.3$$

The sample arithmetic mean for As is 66.3 mg/kg.

Geometric mean

$$\text{sample geometric mean} = \sqrt[n]{(X_1 \times X_2 \times \dots X_n)}$$

$$\text{sample geometric mean} = \sqrt[16]{(103 \times 50 \times \dots 13)}$$

$$\text{sample geometric mean} = 35.2$$

The sample geometric mean for As is 35.2 mg/kg.

As Table 2 shows, each method provides a different result for the measure of central tendency.

Table 2   Variation in central tendency by method of calculation

Method	Result (mg/kg)
Median	38.5
Arithmetic mean	66.3
Geometric mean	35.2

For sample data that is skewed, as in this case, the median and geometric mean are similar, while the arithmetic mean is 'dragged' to the right because of the outliers in the dataset. For a nearly-normal dataset, the three measures would be similar.

The appropriate measure of central tendency should be chosen to represent the sampling data according to the contaminant distribution and the proposed use of the selected measure.

Appendix D: Determining measures of variability

An important aspect of data analysis is determining the variability of the data. Calculating variability can indicate how heterogenous a contaminant is likely to be across a site. The variability of the data is measured by variance, standard deviation or the coefficient of variation.

Variance, represented by s^2, is the average squared distance of each data point from the sample mean, and can be affected by extreme values and large numbers of values below the limits of reporting (LORs). It is used to estimate the population variance σ^2.

The **standard deviation** of a sample, represented by **s**, is calculated by taking the square root of the variance, and indicates the population's typical deviation from the mean. The standard deviation of the population, represented by **σ**, is generally unknown when assessing site contamination, so **s** is used as an estimate. Note that although s^2 is an unbiased estimate of σ^2, s is a **biased** estimate of σ.

The **coefficient of variation** (CV) or **relative standard deviation** (RSD) measures the relative homogeneity of a distribution. The CV is the standard deviation of a distribution divided by the mean of the distribution. The RSD is determined in the same way but expressed as a percentage.

Determination

The methods for determining the measures of variability are shown below.

Variance

$$s^2 = \frac{\Sigma(x_i - \overline{x})^2}{n - 1}$$

Standard deviation of a sample

$$s = \sqrt{\frac{\Sigma(x_i - \overline{x})^2}{n - 1}}$$

Estimate of standard deviation

Where sampling data are not available, an estimate of the standard deviation can be made by dividing the expected range by six, that is, three standard deviations in each direction, as this should represent approximately 99.7% of a nearly-normal distribution.

$$\sigma_E = \frac{C_H - C_L}{6}$$

The relative standard deviation is determined in the same way, but expressed as a percentage, that is, a CV of 0.5 = an RSD of 50%.

Coefficient of variation

$$CV = \frac{s}{\overline{x}}$$

Where:

s^2	variance
x_i	the value of the sample
$\overline{x}$	the arithmetic mean (see Appendix C)
n	number of samples
s	standard deviation

σ_E estimate of population standard deviation

C_H estimate of the highest possible value in the sampling area

C_L estimate of the lowest possible value in the sampling area

CV   coefficient of variation

RSD   relative standard deviation

Worked example

In this example we determine the measures of variability for the sampling data for arsenic (As) in Table 1.

The values for As, shown in mg/kg, are:103, 50, 43, 9, 203, 54, 341, 34, 71, 14, 62, 6, 27, 7, 24 and 13.

The number of samples, n, is 16, and the arithmetic average of the sampling data is 66.3.

Variance

$$s^2 = \frac{\Sigma(x_i - \overline{x})^2}{n - 1}$$

$$s^2 = \frac{(103 - 66.3)^2 + (50 - 66.3)^2 \dots (13 - 66.3)^2}{16 - 1}$$

$$s^2 = \frac{1{,}346.9 + 265.7 + \cdots 2{,}840.9}{15}$$

$$s^2 = 7{,}792.2$$

Standard deviation

$$s = \sqrt{\frac{\Sigma(x_i - \overline{x})^2}{n-1}}$$

$$s = \sqrt{7{,}792.2}$$

$$s = 88.3$$

Estimate of standard deviation

$$\sigma_E = \frac{C_H - C_L}{6}$$

$$\sigma_E = \frac{341 - 6}{6}$$

$$\sigma_E = 55.8$$

In this example, the standard deviation calculated using the sampling data is much greater than the estimate of the standard deviation. This is because the sampling data is skewed to the right and does not follow a nearly-normal distribution.

This example shows that, while estimates of standard deviation can be determined when sampling data are not available, they should always be used with caution. If required sample numbers were calculated using an estimated value such as the one in this example, the result would be too low. Accordingly, the sampling data should be used to refine the assumptions made as part of systematic planning.

Coefficient of variation (CV)

$$CV = \frac{s}{\bar{x}}$$

$$CV = \frac{88.3}{66.3}$$

$$CV = 1.3$$

Relative standard deviation (RSD)

$$RSD = 133.1\%$$

In this example, the CV of 1.3 (equivalent to an RSD of 133.1%) shows a distribution not nearly-normal and expected to be skewed to the right. Any statistical inference should assume a log-normal or other non-normal distribution, and use log-normal or non-parametric methods for analysis.

Appendix E: Assessing contaminant distribution

This appendix explains how contaminant distribution can be assessed using commonly available spreadsheet and statistical software, as discussed in Section 2.7.. The sampling data in this example is sourced from Table 1.

Table 3   **Graphical presentations of example contamination data**

Figure	Description
Figure 1	Summary statistics: metals in fill (mg/kg) as box-and-whiskers plots showing minimum, first quartile, median, third quartile and maximum
Figure 2	Summary statistics: metals in fill (mg/kg) as box-and-whiskers plots showing minimum, first quartile, median, third quartile and maximum, with adjusted scale
Figure 3	Standardised summary statistics (values/criteria): metals in fill (%) as box-and-whiskers plots showing minimum, first quartile, median, third quartile and maximum
Figure 4	Standardised summary statistics (values/criteria): metals in fill (%) as box-and-whiskers plots showing minimum, first quartile, median, third quartile and maximum, with adjusted scale
Figure 5	Multiple histograms for metals in fill (mg/kg)
Figure 6	Q–Q plot for arsenic (mg/kg)
Figure 7	Q–Q plot for chromium (mg/kg)
Figure 8	Q–Q plot for copper (mg/kg)
Figure 9	Q–Q plot for lead (mg/kg)
Figure 10	Q–Q plot for nickel (mg/kg)
Figure 11	Q–Q plot for zinc (mg/kg).

Figure 1   **Summary statistics, metals in soil (mg/kg) – minimum, first quartile, median, third quartile, maximum**
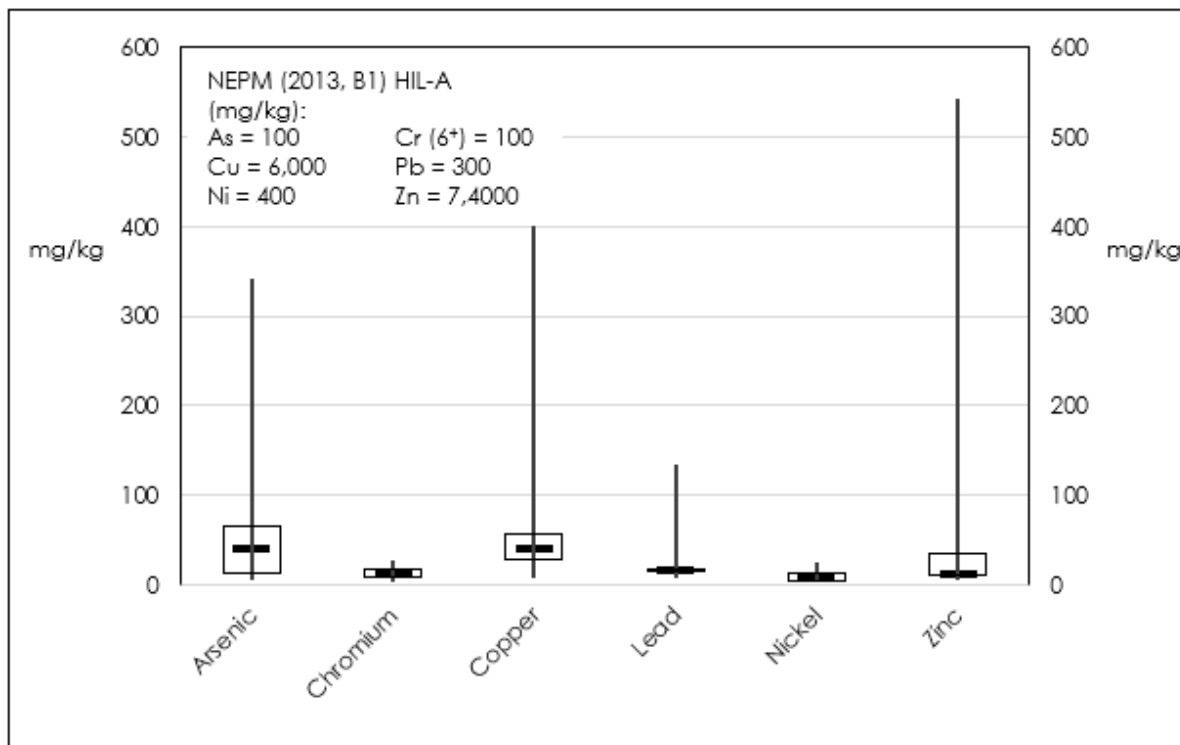
Source: Easterly Point Environmental Pty Ltd

Figure 2   **Summary statistics, metals in soil (mg/kg) – minimum, first quartile, median, third quartile, maximum – scale adjusted**
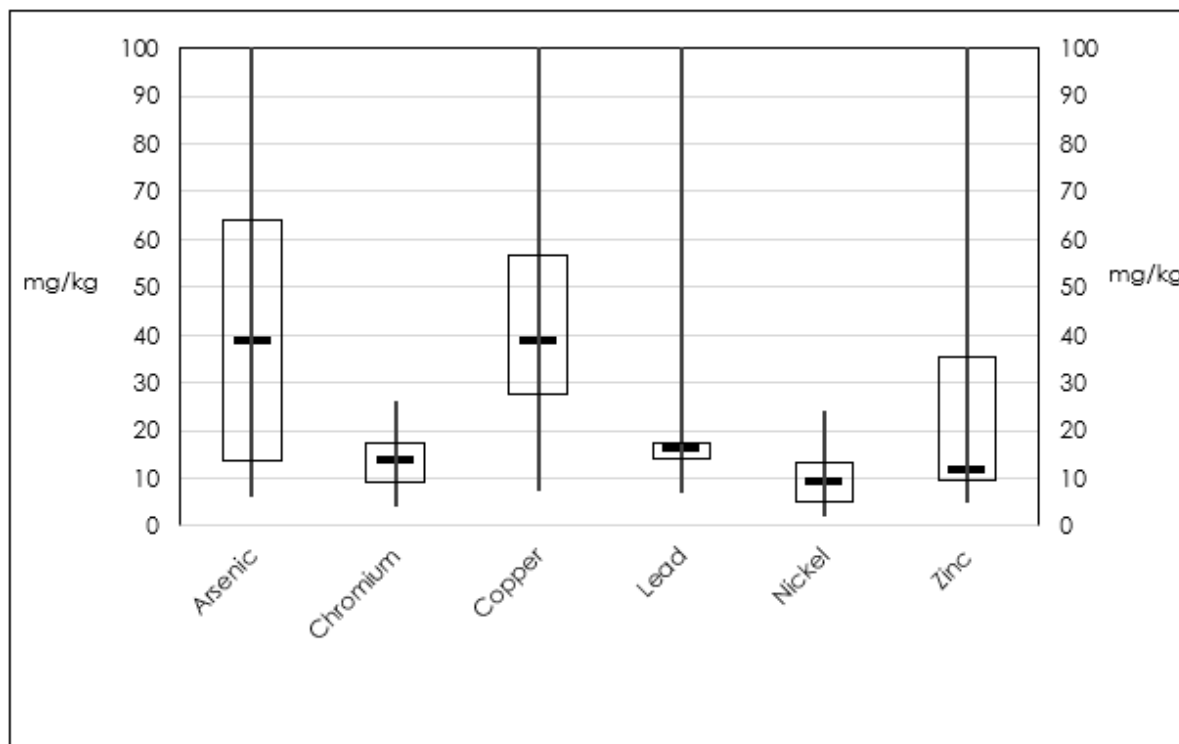
Source: Easterly Point Environmental Pty Ltd

Figure 1 and Figure 2 show the data is generally skewed to the right in the cases of As, Cu, Pb and Zn, as a result of extreme values in the dataset. Cr and Ni look generally symmetrically distributed, suggesting a nearly-normal distribution.

Figure 3   Standardised summary statistics, metals in soil (%) – metals data relative to acceptance criteria
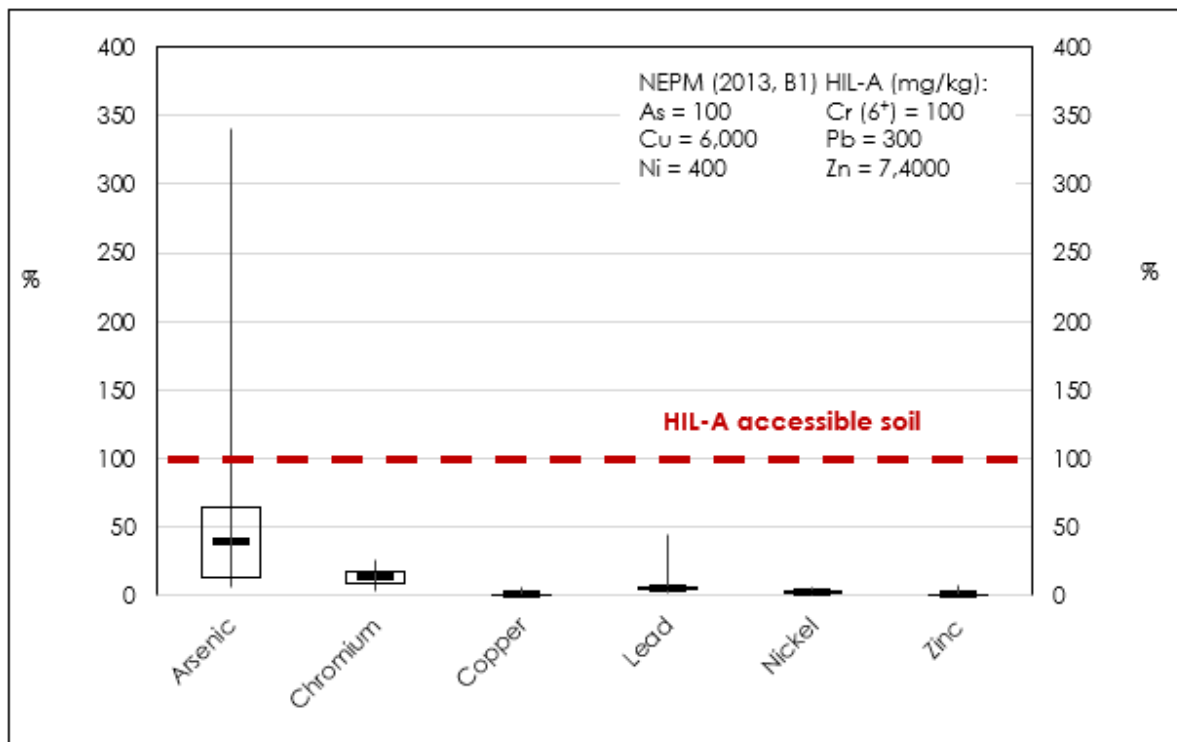
Source: Easterly Point Environmental Pty Ltd

Figure 4   Standardised summary statistics, metals in soil (%) – metals data relative to acceptance criteria – scale adjusted
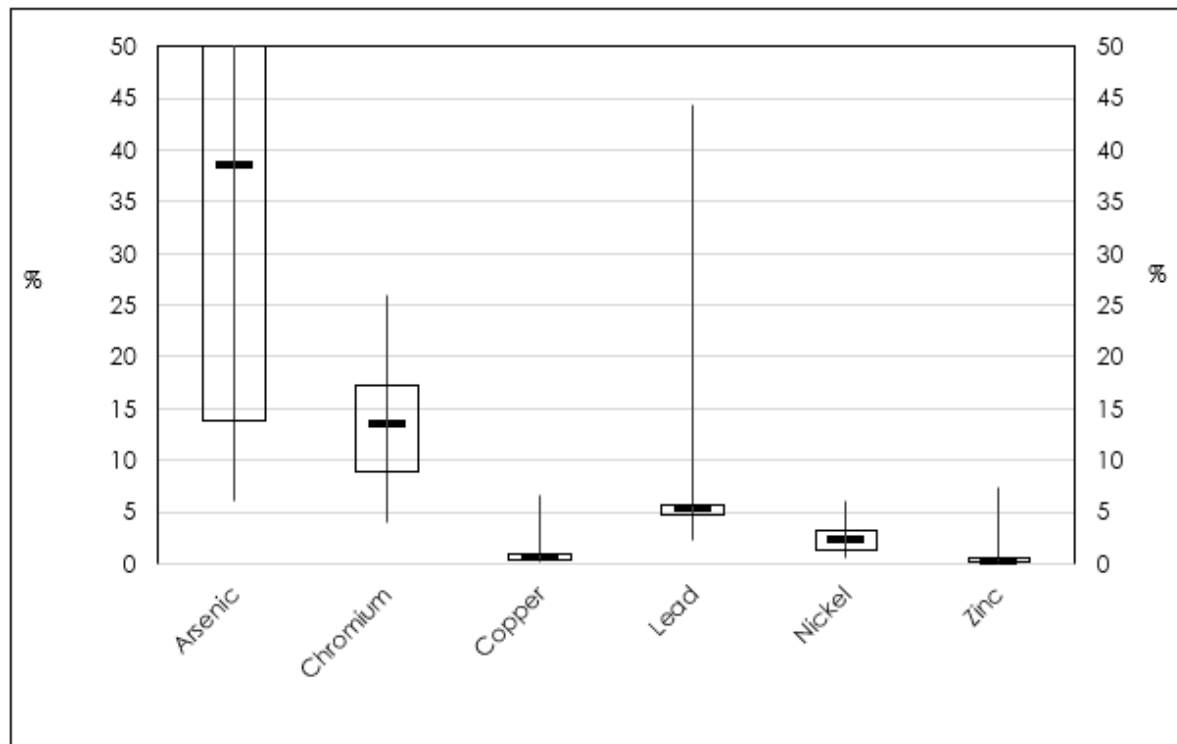
Source: Easterly Point Environmental Pty Ltd

Figures 3 and 4 show that in relation to the criteria for HIL-A residential with accessible soil, only As exceeds 50% of its criterion, with the maximum As value exceeding the criterion by 341%, that is, more than 250% of the criterion. Cu, Pb and Zn are elevated, but are below HIL-A.

Figure 5   Multiple histograms for metals in soil (mg/kg) – data from Table 1
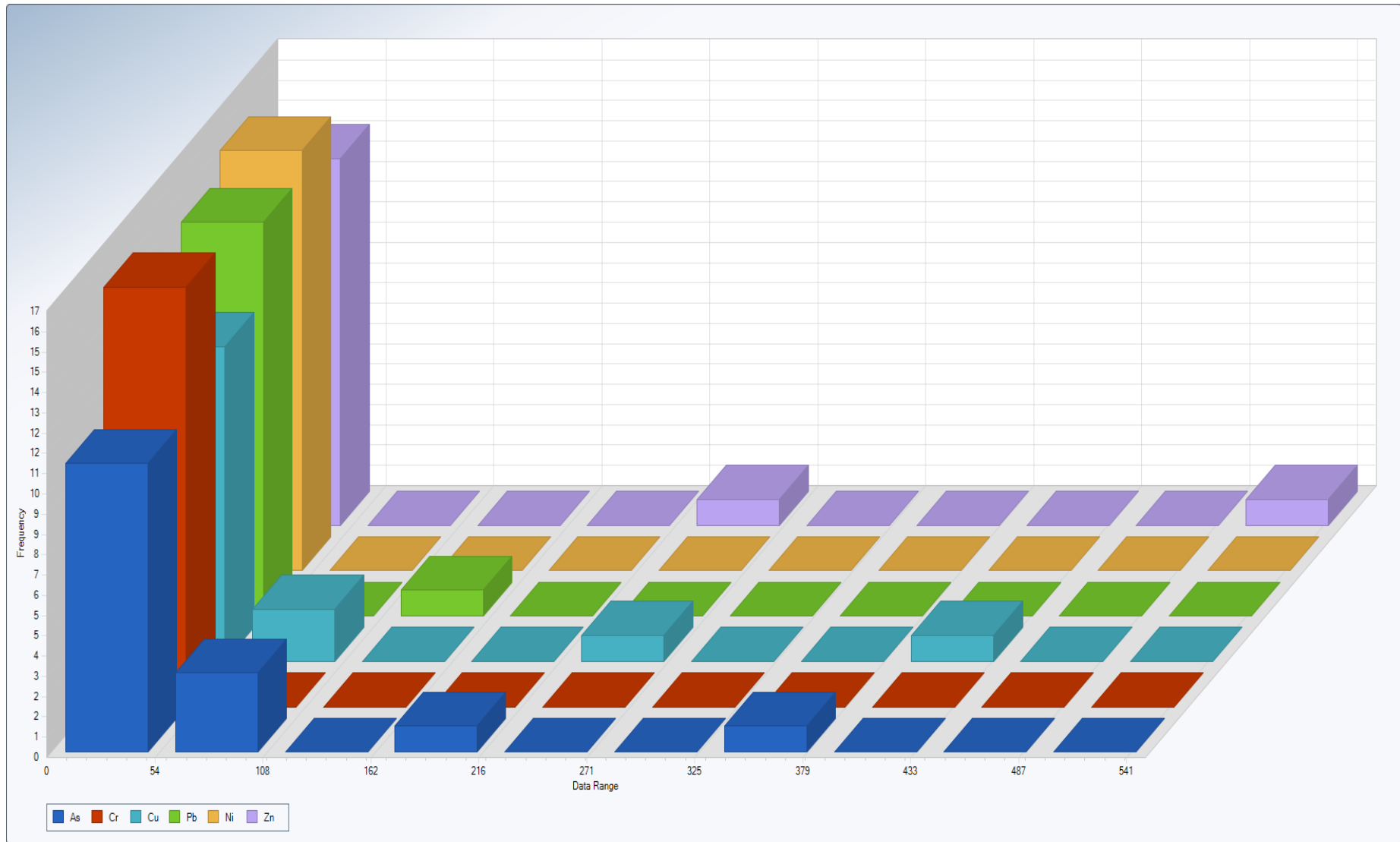
The x-axis shows concentration of the metal (in mg/kg) and the y-axis shows the number of samples.
Outputs from USEPA's ProUCL, created by Easterly Point Environmental Pty Ltd

Figure 5 shows that As, Cu, Pb and Zn are right-skewed because of extreme values. As the sample size was small (less than 30 samples), the normality of the distribution cannot be confirmed using histograms.

Figure 6   Q–Q plot for arsenic (mg/kg)
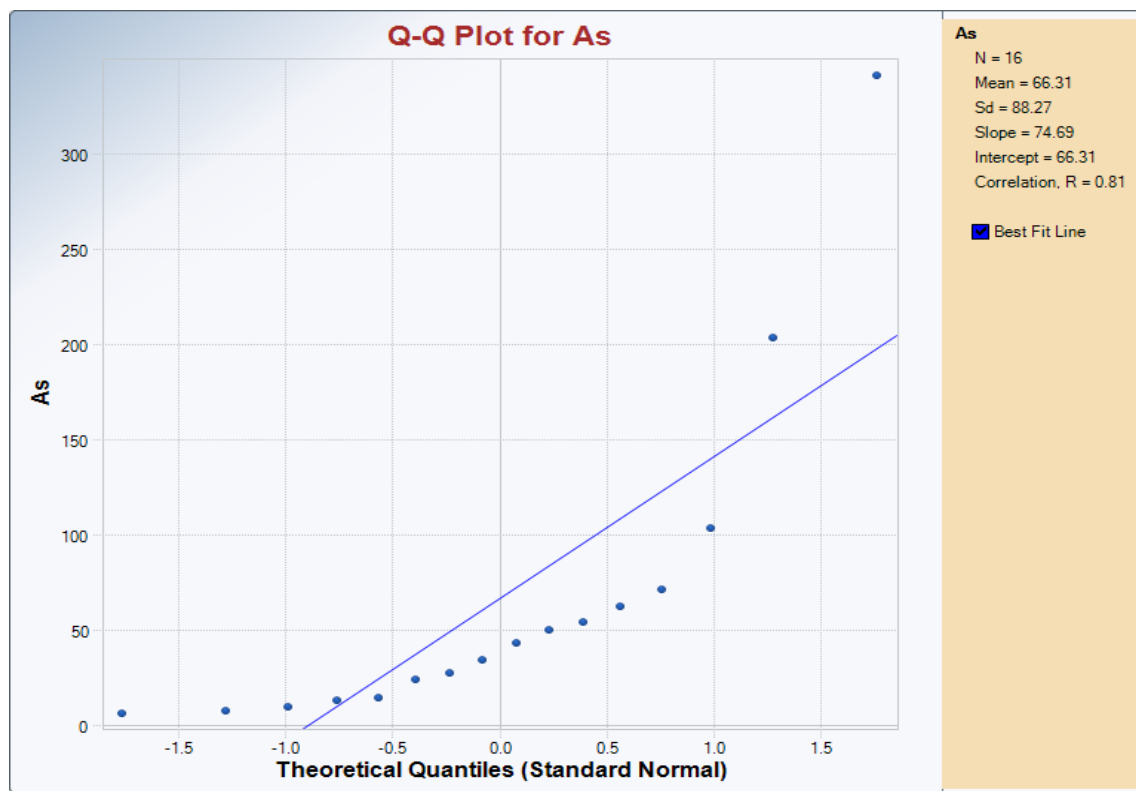
Outputs from USEPA's ProUCL, created by Easterly Point Environmental Pty Ltd

Figure 7   Q–Q plot for chromium (mg/kg)
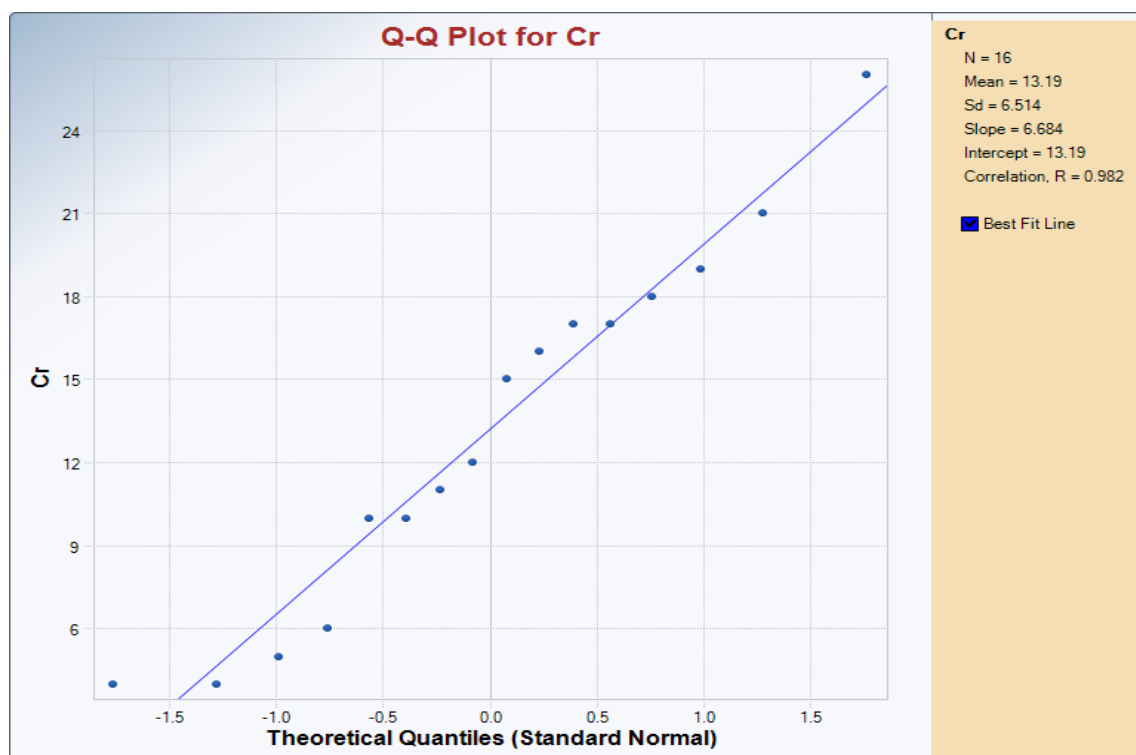
Outputs from USEPA's ProUCL, created by Easterly Point Environmental Pty Ltd

Figure 8   Q–Q plot for copper (mg/kg)
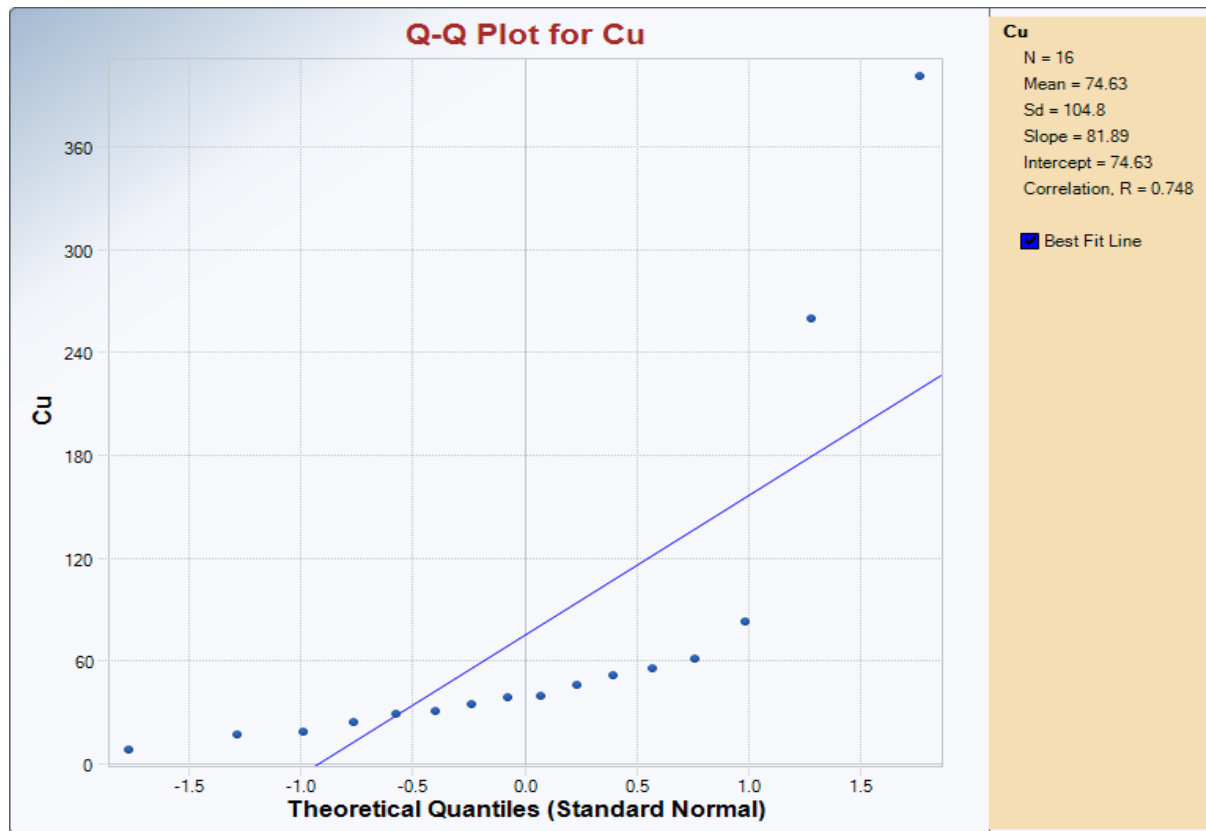
Outputs from USEPA's ProUCL, created by Easterly Point Environmental Pty Ltd

Figure 9   Q–Q plot for lead (mg/kg)
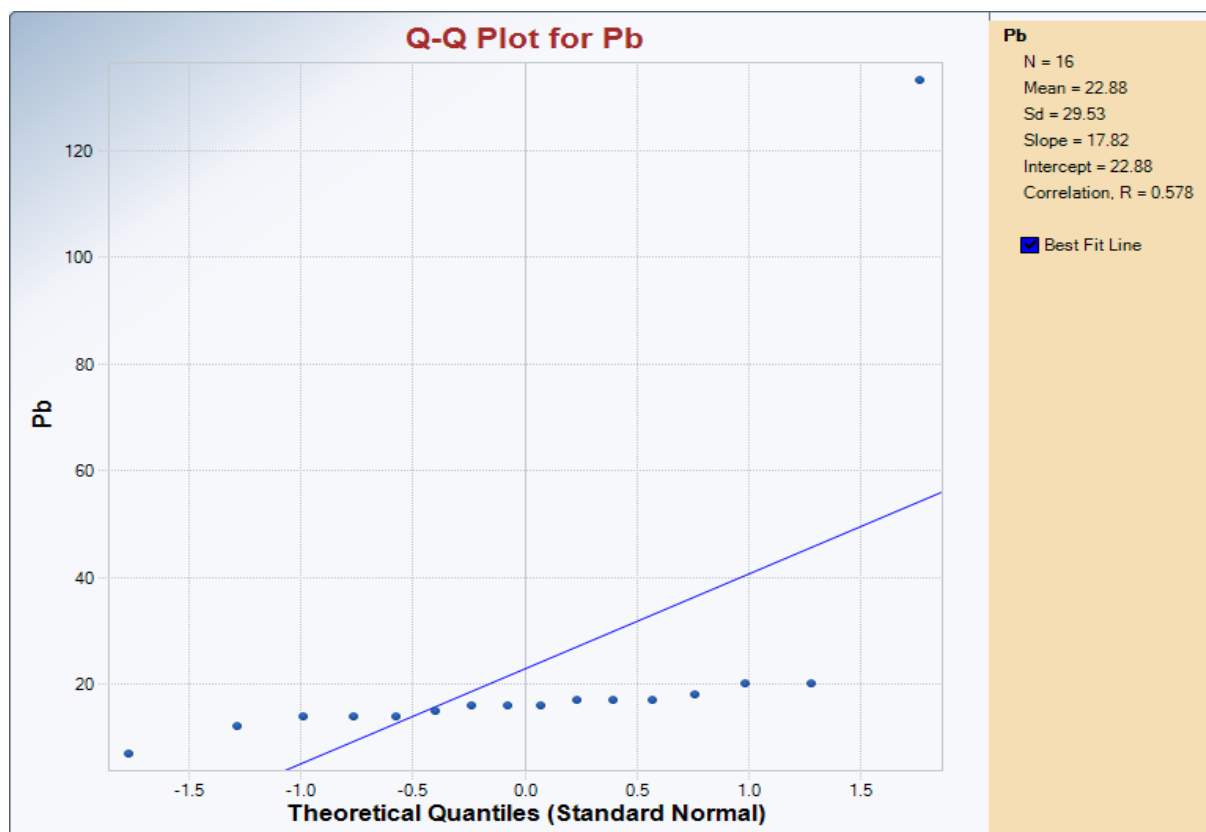
Outputs from USEPA's ProUCL, created by Easterly Point Environmental Pty Ltd

Figure 10   Q–Q plot for nickel (mg/kg)
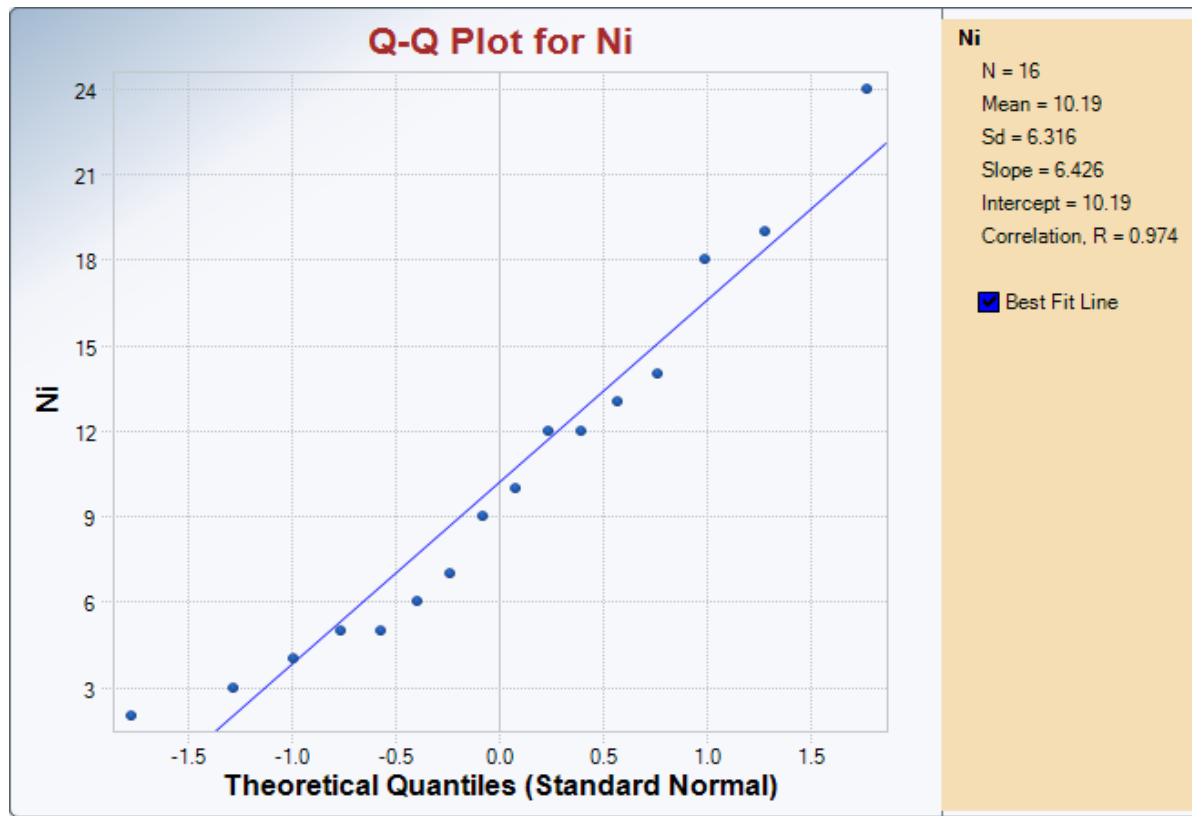
Outputs from USEPA's ProUCL, created by Easterly Point Environmental Pty Ltd

Figure 11   Q–Q plot for zinc (mg/kg)
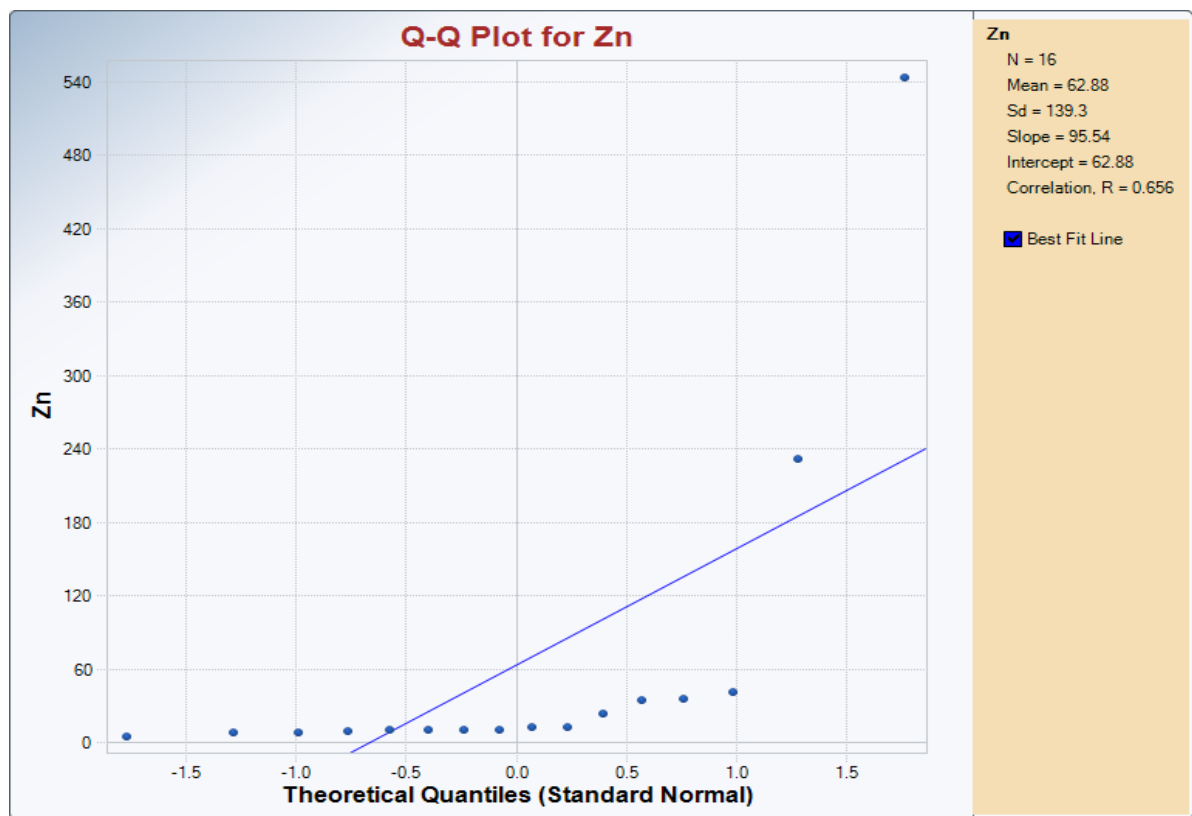
Outputs from USEPA's ProUCL, created by Easterly Point Environmental Pty Ltd

The Q–Q plots in Figures 6–11 show that As, Cu, Pb and Zn are unlikely to be nearly-normally distributed, so parametric methods that assume near-normality cannot be used for statistical inference.

Instead, some form of transformation or another distribution type should be used. For Cr and Ni, the Q–Q plots suggest a nearly-normal distribution, so parametric methods that assume near-normality may be appropriate for analysis.

Appendix F: One-sample t-test hypothesis testing

In the assessment of contaminated land, when a decision requires comparing a sampled population to a target value, such as a specified health investigation level (HIL), a one-sample t-test can be used. This is a parametric method, and assumes a nearly normal distribution, at least for sample sizes of less than 30: it is not suitable for highly skewed datasets. See USEPA 2006a, G-9S for non-parametric methods.

Determination

Establish the null hypothesis (H_0) and alternative hypothesis (H_A). EPA policy is to always assume that the site or decision area is contaminated, so the null hypothesis is always written as:

$$H_0: \mu > \text{criterion or action level}$$

The alternative hypothesis for a one-sided test is then:

$$H_A: \mu \leq \text{criterion or action level}$$

The test statistic (t_0) is calculated using the t-score formula:

$$t_0 = \frac{\bar{x} - C}{s/\sqrt{n}}$$

Where:

μ	population mean
t_0	test statistic
$\bar{x}$	sample mean
C	criterion or action level
s	sample standard deviation
n	number of samples
$t\alpha$	critical value.

The critical value ($t\alpha$) is determined from a table of critical values of Student's t-distribution (see Table 4) or by using an appropriate software program. The confidence level ($1 - \alpha$) and the degrees of freedom ($n - 1$) are used to select $t\alpha$.

The test statistic is then compared to the critical value, and the following decisions made:

- if $t_0 < t\alpha$, then fail to reject the null hypothesis that the true population mean is greater than the criterion or action level
- if $t_0 > t\alpha$, then reject the null hypothesis that the true population mean is greater than the criterion or action level and accept the alternative hypothesis that the true population mean is less than or equal to the criterion or action level.

Whereas the signs of t_0 and $t\alpha$ are important in regard to whether an upper-tailed or lower-tailed test is being conducted, when comparing t_0 to $t\alpha$, it is the absolute values that are compared.

The probability or p-value is also determined, either approximately from a table of critical values of Student's t-distribution (see Table 4) or by using an appropriate software program. This is then compared to the selected value of alpha (α), with the following decisions made:

- if p-value > α, then fail to reject the null hypothesis that the true population mean is greater than the criterion or action level
- if p-value < α, then reject the null hypothesis that the true population mean is greater than the criterion or action level and accept the alternative hypothesis that the true population mean is less than or equal to the criterion or action level.

While the sign of the p-value is important in regard to whether an upper-tailed or lower-tailed test is being conducted, when comparing the p-value to α it is the absolute values that are compared.

As critical values and p-values are mathematically related, either approach will always provide the same conclusion.

Worked example

In this example we use the arsenic (As) and lead (Pb) data from Table 1 to determine whether the null hypothesis (H_0) should be rejected in favour of the alternative hypothesis (H_A). The selected criteria are the HILs for a residential land use (HILs-A), and the test is to be conducted at a confidence level of 95%, that is $\alpha = 0.05$.

The null hypothesis is:

$$H_0: \mu > \text{criterion}$$

The alternative hypothesis is then:

$$H_A: \mu \leq \text{criterion}.$$

The test statistic (t_0) is calculated using the t-score formula:

$$t_0 = \frac{\overline{x} - C}{s/\sqrt{n}}$$

For As, n = 16, $\overline{x}$ = 66.3, s = 88.3 and HIL-A = 100, so:

$$t_0 = \frac{66.3 - 100}{88.3/\sqrt{16}}$$

$$t_0 = -1.53$$

For Pb, n = 16, $\bar{x}$ = 22.9, s = 29.5 and HIL-A = 300, therefore:

$$t_0 = \frac{22.9 - 300}{29.5 / \sqrt{16}}$$

$$t_0 = -37.54$$

From a table of critical values of Student's t-distribution (see Table 4), at a confidence level of 95% for 15 degrees of freedom, $t\alpha = 1.75$.

Critical value

For As, as 1.53 is less than 1.75, that is, $t_0 < t\alpha$, fail to reject the null hypothesis that the true population mean is greater than the criterion.

For Pb, as 37.54 is more than 1.75, that is, $t_0 > t\alpha$, reject the null hypothesis that the true population mean is greater than the criterion and accept the alternative hypothesis that the true population mean is less than or equal to the criterion.

P-value

For As, from a table of critical values of Student's t-distribution (see Table 4), the p-value is between 0.1 and 0.05, that is, $t\alpha$ is between 1.34 and 1.75. Using a software package, the p-value is calculated to be 0.074. As 0.074 is more than 0.05, that is, the p-value $> \alpha$, fail to reject the null hypothesis that the true population mean is greater than the criterion.

For Pb, from a table of critical values of Student's t-distribution, the p-value is less than 0.005, that is, $t\alpha$ is > 2.95. Using a software package, the p-value is calculated to be 1.5×10^{-16}. As 1.5×10^{-16} is less than 0.05, that is, the p-value $< \alpha$, reject the null hypothesis that the true population mean is greater than the criterion and accept the alternative hypothesis that the true population mean is less than or equal to the criterion.

Critical region

In the case of As, t_0 does not fall within the critical region (the area beyond the critical value, $t\alpha$). It is therefore unlikely that the observed test statistic is more extreme than would be expected if the null hypothesis were true. Similarly, as the p-value $> \alpha$, the probability of observing a p-value as extreme as 0.074 would be high, if H_0 were true. Based on both the critical value approach and the p-value approach, there is insufficient evidence at a 95% confidence level to conclude that the population mean for As is less than HIL-A.

In the case of Pb, t_0 falls within the critical region, and it is likely that the observed test statistic is more extreme than would be expected if the null hypothesis were true. And, as the p-value $< \alpha$, the probability of observing a p-value as extreme as 1.5×10^{-16} would be low, if H_0 were true. Based on both the critical value approach and the p-value approach, there is sufficient evidence at a 95% confidence level to reject the null hypothesis and to accept the alternative hypothesis that the population mean for Pb is less than HIL-A.

Table 4   Critical values of the Student's t-distribution

Degrees of freedom	Significance level for one-sided interval (α), e.g. confidence limits	15%	10%	5%	2.5%	1%	0.5%
	Confidence level for one-sided interval ($t_{1-\alpha}$), e.g. confidence limits	85%	90%	95%	97.5%	99%	99.5%
	Significance level for two-sided interval ($\alpha/2$), e.g. confidence intervals	30%	20%	10%	5%	2%	1%
	Confidence level for two-sided interval ($t_{1-\alpha/2}$), e.g. confidence intervals	70%	80%	90%	95%	98%	99%
1		1.963	3.078	6.314	12.706	31.821	63.657
2		1.386	1.886	2.920	4.303	6.965	9.925
3		1.250	1.638	2.353	3.182	4.541	5.841
4		1.190	1.533	2.132	2.776	3.747	4.604
5		1.156	1.476	2.015	2.571	3.365	4.032
6		1.134	1.440	1.943	2.447	3.143	3.707
7		1.119	1.415	1.895	2.365	2.998	3.499
8		1.108	1.397	1.860	2.306	2.896	3.355
9		1.100	1.383	1.833	2.262	2.821	3.250
10		1.093	1.372	1.812	2.228	2.764	3.169
11		1.088	1.363	1.796	2.201	2.718	3.106
12		1.083	1.356	1.782	2.179	2.681	3.055
13		1.079	1.350	1.771	2.160	2.650	3.012
14		1.076	1.345	1.761	2.145	2.624	2.977
15		1.074	1.341	1.753	2.131	2.602	2.947
16		1.071	1.337	1.746	2.120	2.583	2.921
17		1.069	1.333	1.740	2.110	2.567	2.898
18		1.067	1.330	1.734	2.101	2.552	2.878
19		1.066	1.328	1.729	2.093	2.539	2.861
20		1.064	1.325	1.725	2.086	2.528	2.845
21		1.063	1.323	1.721	2.080	2.518	2.831
22		1.061	1.321	1.717	2.074	2.508	2.819
23		1.060	1.319	1.714	2.069	2.500	2.807
24		1.059	1.318	1.711	2.064	2.492	2.797
25		1.058	1.316	1.708	2.060	2.485	2.787
26		1.058	1.315	1.706	2.056	2.479	2.779
27		1.057	1.314	1.703	2.052	2.473	2.771
28		1.056	1.313	1.701	2.048	2.467	2.763
29		1.055	1.311	1.699	2.045	2.462	2.756

Degrees of freedom	Significance level for one-sided interval (α), e.g. confidence limits	15%	10%	5%	2.5%	1%	0.5%
	Confidence level for one-sided interval ($t_{1-\alpha}$), e.g. confidence limits	85%	90%	95%	97.5%	99%	99.5%
	Significance level for two-sided interval ($\alpha/2$), e.g. confidence intervals	30%	20%	10%	5%	2%	1%
	Confidence level for two-sided interval ($t_{1-\alpha/2}$), e.g. confidence intervals	70%	80%	90%	95%	98%	99%
30		1.055	1.310	1.697	2.042	2.457	2.750
40		1.050	1.303	1.684	2.021	2.423	2.704
60		1.046	1.296	1.671	2.000	2.390	2.660
120		1.041	1.289	1.658	1.980	2.358	2.617
∞		1.036	1.282	1.645	1.960	2.326	2.576

Modified from USEPA 2006a G-9S.

Appendix G: Two-sample t-test hypothesis testing

When assessing contaminated land, a decision may require two independent populations to be compared – for example, a potentially contaminated area and a background area, or concentration levels from up-gradient monitoring wells and downgradient monitoring wells. In such cases, a two-sample t-test can be used.

This is a parametric method, so the assumption of normality should be checked; see USEPA 2006a, G-9S for non-parametric methods, if those are required. Two-sample t-tests can also be used for paired populations, such as concentrations before and after remediation; again, see USEPA 2006a, G-9S for parametric and non-parametric methods for paired data.

The method used for conducting a two-sample t-test varies depending on whether the variances (s^2) of the two samples are equal or unequal. For environmental data, the variances are generally unequal, and this method is used in the following determination.

Determination

Establish the null hypothesis (H_0) and alternative hypothesis (H_A). As the objective is to compare two populations, the null hypothesis is set to be that the two populations are equal:

$$H_0: \mu_1 - \mu_2 = \delta_0$$

The alternative hypothesis for a one-sided test is then:

$$H_A: \mu_1 - \mu_2 > \delta_0$$

To calculate the test statistics (t_0) for unequal variance, it is first necessary to determine the degrees of freedom (df) using the Welch–Satterthwaite equation:

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(s_1^2\right)^2}{n_1^2(n_1\text{-}1)} + \frac{\left(s_2^2\right)^2}{n_2^2(n_2\text{-}1)}}$$

The test statistic, t_0, is then calculated using the Welch's t-test formula, which a modification of the Student's t-test formula:

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Where:

μ_1 population 1

μ_2 population 2

df   degrees of freedom

s_1^2 sample variance from population 1

s_2^2 sample variance from population 2

n_1 number of samples from population 1

n_2 number of samples from population 2

t_0 test statistic

$t\alpha$ critical value

$\bar{x}_1$ sample mean from population 1

$\bar{x}_2$ sample mean from population 2

δ_0 difference (delta) of zero

Critical value

The critical value ($t\alpha$) is determined from a table of critical values of Student's t-distribution (see Table 4) or using an appropriate software program. The confidence level ($1 - \alpha$) and the degrees of freedom are used to select $t\alpha$.

The test statistic is then compared to the critical value, and the following decisions made:

- if $t_0 < t\alpha$, fail to reject the null hypothesis that the difference between the population means is zero
- if $t_0 > t\alpha$, reject the null hypothesis that the difference between the population means is zero and accept the alternative hypothesis that the mean of population 1 is greater than the mean of population 2.

While the signs of t_0 and $t\alpha$ are important regarding whether an upper-tailed or lower-tailed test is being conducted, when comparing t_0 to $t\alpha$, it is the absolute values that are compared.

p-value

The probability or p-value is also determined, either approximately from a table of critical values of Student's t-distribution (see Table 4) or using an appropriate software program. The p-value is then compared to the selected value of alpha (α) and the following decisions made:

- if p-value > α, fail to reject the null hypothesis that the difference between the population means is zero
- if p-value < α, reject the null hypothesis that the difference between the population means is zero and accept the alternative hypothesis that the mean of population 1 is greater than the mean of population 2.

While the sign of the p-value is important regarding whether an upper-tailed or lower-tailed test is being conducted, when comparing the p-value to α, it is the absolute values that are compared.

As critical values and p-values are mathematically related, either approach will always provide the same conclusion.

Worked example

In this example we use the arsenic (As) data from Table 1 to determine whether contamination is limited only to the surficial soils (population 1), and therefore if the deeper soils (population 2) can be considered separately. The descriptive statistics for the two datasets, and the original combined dataset for comparison, are shown in Table 5.

Table 5   Arsenic summary statistics by population (mg/kg) – simulated data from Table 1

Statistic	Surface population 1	Depth population 2	Combined
Maximum	341	54	341
Mean	109.3	23.4	66.3
Medium	66.5	13.5	38.5
Minimum	24	6	6
Variance	12,093.4	390.3	7,792.2
Standard deviation	110.0	19.8	88.3

The test is to be conducted at a confidence level of 95%, i.e. $\alpha = 0.05$.

The null hypothesis is:

$$H_0: \mu_1 - \mu_2 = \delta_0$$

The alternative hypothesis for a one-sided test is then:

$$H_A: \mu_1 - \mu_2 > \delta_0$$

The degrees of freedom is first calculated using the Welch–Satterthwaite equation:

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(s_1^2\right)^2}{n_1^2(n_1 - 1)} + \frac{\left(s_2^2\right)^2}{n_2^2(n_2 - 1)}}$$

$$df = \frac{\left(\frac{12,093.4}{8} + \frac{390.3}{8}\right)^2}{\frac{(12,093.4)^2}{8^2(8 - 1)} + \frac{(390.3)^2}{8^2(8 - 1)}}$$

$$df = \frac{1,560.5^2}{3.3 \times 10^5 + 340}$$

$$df = 7.45$$

Rounded down to the next integer, the degrees of freedom is seven (7). A conservative approach is to estimate the degrees of freedom by using the smaller of $n_1 - 1$ or $n_2 - 1$: in this case, that number is also seven.

The test statistic, t_0, is then calculated using Welch's t-test formula:

$$t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - \delta_0}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

$$t_0 = \frac{(109.3 - 23.4) - 0}{\sqrt{\left(\dfrac{12{,}093.4}{8}\right) + \dfrac{390.3}{8}}}$$

$$t_0 = 2.174$$

Critical value

From a table of critical values of Student's t-distribution (see Table 4), at a confidence level of 95% for seven degrees of freedom, $t\alpha = 1.895$.

As 2.174 is more than 1.895, that is, $t_0 > t\alpha$, the null hypothesis that the population means are equal is rejected, and the alternative hypothesis H_A (that the mean of population 1 is greater than the mean of population 2) is accepted, i.e. $\mu_1 - \mu_2 > \delta_0$.

p-value

From a table of critical values of Student's t-distribution, the p-value is between 0.025 and 0.05, that is, $t\alpha$ is between 2.365 and 1.895. Using a software package, the p-value is calculated to be 0.033.
As 0.033 is less than 0.05, that is, the p-value $< \alpha$, the null hypothesis H_0 (that the population means are equal) is rejected, and the alternative hypothesis H_A (that the mean of population 1 is greater than the mean of population 2) is accepted, i.e. $\mu_1 - \mu_2 > \delta_0$.

Critical region

As t_0 falls within the critical region, it is likely that the observed test statistic is more extreme than would be expected if the null hypothesis were true. And, as the p-value $< \alpha$, the probability of observing a p-value as extreme as 0.033 would be low, if H_0 were true. Both the critical-value approach and the p-value approach give sufficient evidence at a 95% confidence level to reject the null hypothesis and accept the alternative hypothesis that the mean of population 1 is greater than the mean of population 2.

Based on review of the summary data, relative to a HIL-A of 100 mg/kg, and the results of the two-sample t-test, it appears that significant impacts relate to the surficial soils rather than the deeper soils. Accordingly, for the design of further investigations and consideration of remedial options, the surficial soils and deeper soils should be considered as separate decision areas. The actual depths which these two populations encompass will need to be determined by further investigations.

Appendix H: Decision errors

Statistical hypothesis testing using a null hypothesis significance testing (NHST) framework – the testing of the null hypothesis (H_0) against an alternative hypothesis (H_A) – can lead to the following four outcomes:

- accepting H_0 when H_0 is true – this is a correct decision for the confidence level of the test $(1 - \alpha)$, for example, $\alpha = 0.05$ and confidence level $= 95\%$
- rejecting H_0 when H_0 is true – this is a Type I or α decision error and results in the false rejection of H_0
- accepting H_0 when H_0 is false – this is a Type II or β decision error and results in the false acceptance of H_0
- rejecting H_0 when H_0 is false – this is a correct decision – the power of the test is $(1 - \beta)$, for example, $\beta = 0.20$ and power $= 80\%$.

These outcomes are summarised in Table 6.

Table 6   **Decision errors in hypothesis testing**

Decision made	Actual condition – H_0 is true	Actual condition – H_0 is false (H_A is true)
Accept H_0 (fail to reject H_0)	Correct decision $1 - \alpha$ = confidence level	Decision error (Type II error) **False acceptance**
Reject H_0 (accept H_A)	Decision error (Type 1 error) **False rejection**	Correct decision $1 - \beta$ = power of test

In this instance, the null hypothesis is that the site or decision area is contaminated.

Decision errors are therefore generally defined as follows:

- The site or decision area is considered not to be contaminated when it actually is – a Type I error. Type I errors can lead to unacceptable risks to human health and the environment, so the regulatory framework is established to protect against Type I errors.
- The site or decision area is considered to be contaminated when it actually is not – a Type II error. Type II errors can lead to sites or decision areas being remediated unnecessarily, or land being used for a less-sensitive land use, or unwarranted restrictions on the surrounding environment, such as water-use restrictions or fishing bans.

Appendix I: 95% confidence intervals

Confidence intervals can be used as an indicator of uncertainty around a point estimate, in this case the mean. By choosing a method for expressing uncertainty, a performance metric that quantifies uncertainty can be specified, allowing limits to be established against which the quality and quantity of the data can be compared (USEPA 2006b, G-4).

A method for determining the 95% confidence interval (CI) of the mean for a nearly-normal distribution is shown in this appendix, using the Student's t formula. For mildly skewed datasets, the Student's t-statistic should be used, but for moderate to highly skewed datasets, the confidence interval based on the t-statistic can fail to cover the population mean, especially for small sample sizes (USEPA 2006a, G-9S). It is therefore important to test the data for normality. This is most easily done by constructing normal Q–Q plots, using appropriate statistical software packages. For other distributions or non-parametric methods, refer to USEPA 2006a, G-9S.

Determination

The test statistic is calculated using the two-sided Student's t-UCL formula:

$$95\% \text{ confidence interval} = \bar{x} \pm t_{\alpha/2,\, n-1} \frac{s}{\sqrt{n}}$$

Where:

$\bar{x}$ sample mean

$t_{\alpha/2, n-1}$ critical value

s sample standard deviation

n number of samples

$s/\sqrt{n}$ standard error of the mean ($SE\bar{x}$).

The standard error of the mean ($SE\bar{x}$) describes the variability in the sampling distribution, that is, the distribution of means from multiple sampling events of the same population, not the variability in the underlying population. One key feature of the $SE\bar{x}$ is that it decreases as the sample size increases (Devore and Farnum 2005).

The $SE\bar{x}$ multiplied by the critical value gives the margin of error (MoE), which can be defined as the radius, or half the width, of a confidence interval for a particular statistic at a specified confidence level (in the equation above, at a 95% confidence level). The MoE also decreases as the number of samples increases.

The critical value is determined from a table of critical values of Student's t-distribution (Table 4 in Appendix F) or using an appropriate statistical software package. The confidence level $(1 - \alpha)$ and the degrees of freedom $(n - 1)$ are used to select $t_{\alpha/2, n-1}$ for a two-sided interval.

Worked example

In this example we use the metals data from Table 1 to determine the 95% confidence interval for chromium (Cr) for surface fill (n = 8) and all fill (n = 16), at a confidence level of 95% ($\alpha = 0.05$).

The 95% confidence interval is calculated using the Student's t-UCL formula:

$$95\% \text{ confidence interval} = \bar{x} \pm t_{\alpha/2,\, n-1} \frac{s}{\sqrt{n}}$$

Surface material

The critical value is selected for a two-sided interval from a table of critical values of Student's t-distribution (Table 4 in Appendix F). At seven (7) degrees of freedom the critical value is 2.365.

For surface soil, x̄ = 15.3, s = 6.5 and n = 8:

$$95\% \text{ confidence interval} = 15.3 \pm 2.365 * \frac{6.5}{\sqrt{8}}$$

$$95\% \text{ confidence interval} = 15.3 \pm 5.4$$

$$95\% \text{ confidence interval} = 9.8 \text{ to } 20.7 \text{ mg/kg}$$

All material

The critical value is selected for a two-sided interval from a table of critical values of Student's t-distribution (Table 4 in Appendix F). At 15 degrees of freedom the critical value is 2.131.

For all fill, x̄ = 13.2, s = 6.5 and n = 16:

$$95\% \text{ confidence interval} = 15.3 \pm 2.131 * \frac{6.5}{\sqrt{16}}$$

$$95\% \text{ confidence interval} = 15.3 \pm 3.5$$

$$95\% \text{ confidence interval} = 9.7 \text{ to } 16.7 \text{ mg/kg}$$

Based on similar datasets, the greater number of samples used in the analysis for all soil samples (16) results in a smaller MoE, and therefore a narrower confidence interval, than does the fewer samples used in analysing the surficial soil (8 samples). Figure 12 illustrates this for both Cr and nickel (Ni); Table 7 and Table 8 show the associated summary statistics.

The maximum probable error (MPE), which is a relative measure based on the MoE divided by the mean (MPE = MoE/x̄), can be used to specify the required statistical precision for data collection. For example, for Ni, Table 7 and Table 8 show that the MPE for eight (8) samples is 52.6%, while 16 samples are required to achieve an MPE of 33.0%.

Figure 12   Summary statistics for Cr and Ni data with variable n (mg/kg) – minimum, 95% LCL, mean, 95% UCL, maximum
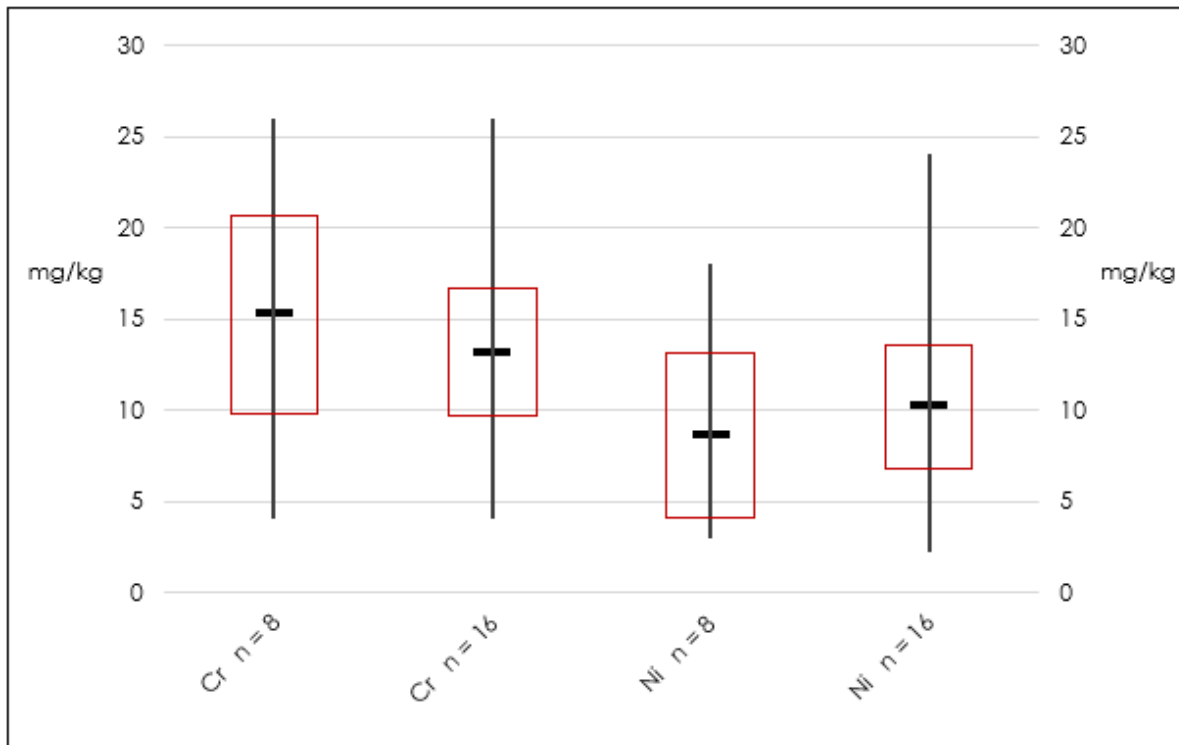
Source: Easterly Point Environmental Pty Ltd

Table 7   Summary statistics for Cr and Ni data (mg/kg) – surface locations

Surface data	Chromium	Nickel
Number of samples	8	8
Sample mean	15.3	8.6
Standard deviation	6.5	5.4
Standard error of the mean (SE$\bar{x}$)	2.3	1.9
Relative standard deviation (RSD)	42.6%	62.9%
Margin of error (MoE)	5.4	4.5
Maximum probable error (MPE)	35.6%	52.6%

Table 8   Summary statistics for Cr and Ni data (mg/kg) – all locations

All data	Chromium	Nickel
Number of samples	16	16
Sample mean	13.2	10.2
Standard deviation	6.5	6.3
Standard error of the mean (SE$\bar{x}$)	1.6	1.6
Relative standard deviation (RSD)	49.4%	61.9%
Margin of error (MoE)	3.5	3.4
Maximum probable error (MPE)	26.3%	33.0%

Appendix J: 95% UCLx̄ for normal distributions

Here we show a method for determining the 95% upper confidence limit of the mean (UCLx̄) for a nearly-normal distribution, using the Student's t formula.

For mildly skewed datasets, the Student's t-statistic should be used, but for moderate to highly skewed datasets, the 95% UCLx̄ based on the t-statistic may not cover the population mean, especially for small sample sizes. It is therefore important to test the data for normality. This is most easily done by constructing normal Q–Q plots, using appropriate statistical software packages.

Determination

The test statistic is calculated using the one-sided Student's t-UCL formula:

$$95\% \text{ UCL}\bar{x} = \bar{x} + t_{\alpha,n-1} \frac{s}{\sqrt{n}}$$

Where:

95% UCLx̄	test statistic
x̄	sample mean
$t_{\alpha,n-1}$	critical value
s	sample standard deviation
n	number of samples

The critical value is determined from a table of critical values of Student's t-distribution (Table 4 in Appendix F), or using an appropriate statistical software package. The confidence level $(1 - \alpha)$ and the degrees of freedom $(n - 1)$ are used to select $t_{\alpha,n-1}$.

Worked example

Here we use the metals data from Table 1 to determine the 95% UCLx̄ for arsenic (As) and chromium (Cr) at a confidence level of 95% ($\alpha = 0.05$), at 15 degrees of freedom (16 − 1 = 15).

The 95% UCLx̄ is calculated using the Student's t-UCL formula:

$$95\% \text{ UCL}\bar{x} = \bar{x} + t_{\alpha,n-1} \frac{s}{\sqrt{n}}$$

The critical value is selected from a table of critical values of Student's t-distribution (Table 4 in Appendix F). In this instance it is 1.753.

Arsenic

For As, x̄ = 66.3, s = 88.3 and n = 16:

$$95\% \text{ UCL}\bar{x} = 66.3 + 1.753 \frac{88.3}{\sqrt{16}}$$

$$95\% \text{ UCL}\bar{x} = 66.3 + 38.7$$

$$95\% \text{ UCL}\bar{x} = 105.0$$

Chromium

For Cr, x̄ = 13.2, s = 6.5 and n = 16:

$$95\% \text{ UCL}\bar{x} = 13.2 + 1.753 \frac{6.5}{\sqrt{16}}$$

$$95\% \ \text{UCL}\overline{x} = 13.2 + 2.85$$

$$95\% \ \text{UCL}\overline{x} = 16.05$$

The coefficient of variation (CV) for As is 1.3, suggesting a distribution that is not nearly-normal: this is confirmed by the Q–Q plot for As (Figure 6 in Appendix E). Figures 1, 2 and 5 show that the dataset is skewed to the right, indicating that it cannot be appropriately analysed with a Student's t-distribution. Running the data through a statistical software package gives the same conclusion: the software recommends the use of a gamma distribution and calculates a 95% UCL$\overline{x}$ of 120.5 mg/kg.

The CV for Cr is 0.5, suggesting a distribution that is nearly-normal: this is confirmed by the distribution shown in Figures 1 and 2 and the Q–Q plot for Cr in Figure 7. Cr appears to be normally and symmetrically distributed, and therefore the calculated value is likely to be an accurate estimate of the 95% UCL$\overline{x}$. A statistical software package confirmed this: it recommended use of a Student's t-distribution and calculated a 95% UCL for the mean of 16.04 mg/kg.

Based on use of the Student's t-UCL formula to calculate these 95% UCL$\overline{x}$, there is a 95% probability that the mean concentration of Cr will not exceed 16.05 mg/kg. The As dataset needs to be analysed further by another method.

Appendix K: 95% UCLx̄ for log-normal distributions

Here we show a method for determining the 95% upper confidence limit of the mean (UCLx̄) for a log-normal distribution, using the Land's H-statistic.

This method assumes log-normality, and it is very important to test this assumption. The easiest way to do this is to construct log-normal Q–Q plots using an appropriate statistical software package.

Determination

The test statistic is calculated using the one-sided Land's H-statistic:

$$95\% \text{ H-UCL}\bar{x} = \exp\left(\bar{y} + \frac{s_y^2}{2} + \frac{s_y H_{1-\alpha}}{\sqrt{n-1}}\right)$$

Where:

95% H-UCLx̄	test statistic
exp	exponential function, that is, 2.7183 to the power of the value inside the brackets
x̄	mean of the log-transformed sample measurements
s_y^2	variance of the log-transformed sample measurements
s_y	standard deviation of the log-transformed sample measurements
$H_{1-\alpha}$	H-statistic critical value, at the stated confidence level $(1 - \alpha)$, which depends on the values of s_y and n
n	number of samples

The sample data is transformed using the natural logarithm, that is, a logarithm to the base **e** (2.7183), so $y_i = \ln x_i$, and the descriptive statistics $\bar{y}$, s_y^2 and s_y are determined from the transformed data.

The value of $H_{1-\alpha}$ is selected from Table 9 for a 95% confidence level, based on the values for s_y and n. For other confidence levels, refer to USEPA 2006a, G-9S, and for values of s_y and n not listed in Table 9, use interpolation.

Worked example

Here we use the metals data from Table 1 to determine the 95% H-UCLx̄ for arsenic (As) and copper (Cu) at a confidence level of 95% ($\alpha = 0.05$).

Arsenic

The sample data is transformed using the natural logarithm, and for As, $\bar{y} = 3.561$, $s_y^2 = 1.347$, $s_y = 1.160$ and n = 16.

The value of H is selected from Table 9. Based on s_y and n, H is between 2.564 and 3.163. By interpolation, H = 2.958.

The test statistic is calculated from:

$$95\% \text{ H-UCL}\overline{x} = \exp\left(\overline{y} + \frac{s_y^2}{2} + \frac{s_y H_{1-\alpha}}{\sqrt{n-1}}\right)$$

$$95\% \text{ H-UCL}\overline{x} = \exp\left(3.561 + \frac{1.347}{2} + \frac{1.160 * 2.958}{\sqrt{16-1}}\right)$$

$$95\% \text{ H-UCL}\overline{x} = \exp(5.120)$$

$$95\% \text{ H-UCL}\overline{x} = 167.4$$

The coefficient of variation (CV) for As is 1.3, suggesting a distribution that is not nearly-normal: this is confirmed by the Q–Q plot for As (Figure 6 in Appendix E). Figures 1, 2 and 5 show that the dataset is skewed to the right. While this suggests that an H-UCL$\overline{x}$ may be appropriate, when a statistical software package was used to generate a range of distributions for calculating the 95% UCL$\overline{x}$, it recommended a gamma distribution.

Although the As data appear log-normal, the Land's H-statistic is sensitive to deviations from log-normality, and produces very high values for large variance or skewness, or where n is small (< 30) (USEPA 2002d). Accordingly, USEPA 2015a recommends that positively skewed datasets should first be tested for a gamma distribution. If the dataset follows a gamma distribution, the UCL$\overline{x}$ should then be computed using a gamma distribution.

Assuming a gamma distribution for the As data, the software package determined a 95% UCL$\overline{x}$ of 120.5 mg/kg – markedly different from the 95% H-UCL$\overline{x}$ of 167.4 mg/kg. As both exceed the HIL-A for As of 100 mg/kg, further data analysis or investigations would be recommended.

Copper

The sample data is transformed using the natural logarithm, and for Cu, $\overline{y} = 3.773$, $s_y^2 = 0.950$, $s_y = 0.974$ and n = 16.

The value of H is selected from Table 9. Based on s_y and n, H is between 2.432 and 2.744. By interpolation, H = 2.619.

The test statistic is calculated from:

$$95\% \text{ H-UCL}\overline{x} = \exp\left(\overline{y} + \frac{s_y^2}{2} + \frac{s_y H_{1-\alpha}}{\sqrt{n-1}}\right)$$

$$95\% \text{ H-UCL}\overline{x} = \exp\left(3.773 + \frac{0.950}{2} + \frac{0.974 * 2.619}{\sqrt{16-1}}\right)$$

$$95\% \text{ H-UCL}\overline{x} = \exp(4.907)$$

$$95\% \text{ H-UCL}\overline{x} = 135.2$$

The CV for Cu is 1.4, suggesting a distribution that is not nearly-normal. This is confirmed by the Q–Q plot for Cu in Figure 8. Figure 1 and Figure 5 show that the dataset is skewed to the right, suggesting that an H-UCL$\overline{x}$ may be appropriate. This was confirmed by using a statistical software package to generate a range of distributions for calculating the 95% UCL$\overline{x}$. In both cases the 95% UCL$\overline{x}$ was 135.2 mg/kg.

Table 9   Values of H for one-sided 95% confidence level for computing H-UCL on a log-normal mean

S_y	n = 3	n = 5	n = 7	n = 10	n = 12	n = 15	n = 21	n = 31	n = 51	n = 101
0.10	2.750	2.035	1.886	1.802	1.775	1.749	1.722	1.701	1.684	1.670
0.20	3.295	2.198	1.992	1.881	1.843	1.809	1.771	1.742	1.718	1.697
0.30	4.109	2.402	2.125	1.977	1.927	1.882	1.833	1.793	1.761	1.733
0.40	5.220	2.651	2.282	2.089	2.026	1.968	1.905	1.856	1.813	1.777
0.50	6.495	2.947	2.465	2.220	2.141	2.068	1.989	1.928	1.876	1.830
0.60	7.807	3.287	2.673	2.368	2.271	2.181	2.085	2.010	1.946	1.891
0.70	9.120	3.662	2.904	2.532	2.414	2.306	2.191	2.102	2.025	1.960
0.80	10.43	4.062	3.155	2.710	2.570	2.443	2.307	2.202	2.112	2.035
0.90	11.74	4.478	3.420	2.902	2.738	2.589	2.432	2.310	2.206	2.117
1.00	13.05	4.905	3.698	3.103	2.915	2.744	2.564	2.423	2.306	2.205
1.25	16.33	6.001	4.426	3.639	3.389	3.163	2.923	2.737	2.580	2.447
1.50	19.60	7.120	5.184	4.207	3.896	3.612	3.311	3.077	2.881	2.713
1.75	22.87	8.250	5.960	4.795	4.422	4.081	3.719	3.437	3.200	2.997
2.00	26.14	9.387	6.747	5.396	4.962	4.564	4.141	3.812	3.533	3.295
2.50	32.69	11.67	8.339	6.621	6.067	5.557	5.013	4.588	4.228	3.920
3.00	39.23	13.97	9.945	7.864	7.191	6.570	5.907	5.388	4.947	4.569
3.50	45.77	16.27	11.56	9.118	8.326	7.596	6.815	6.201	5.681	5.233
4.00	52.31	18.58	13.18	10.38	9.469	8.630	7.731	7.024	6.424	5.908
4.50	58.85	20.88	14.80	11.64	10.62	9.669	8.652	7.854	7.174	6.590
5.00	65.39	23.19	16.43	12.91	11.77	10.71	9.579	8.688	7.929	7.277
6.00	78.47	27.81	19.68	15.45	14.08	12.81	11.44	10.36	9.449	8.661
7.00	91.55	32.43	22.94	18.00	16.39	14.90	13.31	12.05	10.98	10.05
8.00	104.6	37.06	26.20	20.55	18.71	17.01	15.18	13.74	12.51	11.45
9.00	117.7	41.68	29.46	23.10	21.03	19.11	17.05	15.43	14.05	12.85
10.00	130.8	46.31	32.73	25.66	23.35	21.22	18.93	17.13	15.59	14.26

From Gilbert 1987

For values of s_y and n not listed, use interpolation.

For other confidence levels, refer to USEPA 2006a, G-9S.

Appendix L: 95% UCLx̄ for skewed distributions

Here we give a method for determining the 95% upper confidence limit of the mean (UCLx̄) when the distribution cannot be identified. It is based on the non-parametric Chebyshev inequality formula.

The Chebyshev inequality formula makes no assumptions about distribution. For moderately skewed datasets, it yields conservative but realistic values for UCLx̄. For highly skewed datasets, it can substantially underestimate the UCLx̄, especially for small sample sizes, because it assumes that the standard deviation of the underlying distribution is known. In such cases you can use higher confidence limits (USEPA 2015a): statistical software packages will usually recommend these.

Determination

For unknown distributions, the test statistic is calculated using the one-sided Chebyshev inequality formula:

$$95\% \; \text{UCL}\overline{x} = \overline{x} + k_{(1-\alpha)} \; \frac{s}{\sqrt{n}}$$

Where:

95% UCLx̄	test statistic
$\overline{x}$	sample mean
$k_{(1-\alpha)}$	critical value
s	sample standard deviation
n	number of samples

The critical value, k, which is based on the one-sided Chebyshev inequality, is selected from Table 10. It is determined as:

$$k = \sqrt{\frac{1}{\alpha} - 1}$$

Table 10   Critical values based on the Chebyshev theorem

Confidence level %	alpha (α)	k
99	0.01	9.95
95	0.05	4.36
90	0.10	3.00
85	0.15	2.38
80	0.20	2.00
75	0.25	1.73

Adapted from CL:AIRE (2008).

Worked example

Here the metals data from Table 1 is used to determine the 95% UCL of the mean for arsenic (As) and zinc (Zn), at a confidence level of 95% ($\alpha = 0.05$).

The test statistic is calculated using the Chebyshev inequality formula:

$$95\% \; \text{UCL}\overline{x} = \overline{x} + k_{(1-\alpha)} \; \frac{s}{\sqrt{n}}$$

The critical value is selected from Table 10. For $\alpha = 0.05$, $k_{(1-\alpha)} = 4.36$.

Arsenic

For As, $\bar{x} = 66.3$, $s = 88.3$ and $n = 16$:

$$95\% \text{ UCL}\bar{x} = 66.3 + 4.36\frac{88.3}{\sqrt{16}}$$

$$95\% \text{ UCL}\bar{x} = 66.3 + 96.2$$

$$95\% \text{ UCL}\bar{x} = 162.5$$

Zinc

For Zn, with $\bar{x} = 62.9$, $s = 139.3$ and $n = 16$:

$$95\% \text{ UCL}\bar{x} = 62.9 + 4.36\frac{139.3}{\sqrt{16}}$$

$$95\% \text{ UCL}\bar{x} = 62.9 + 151.9$$

$$95\% \text{ UCL}\bar{x} = 214.7$$

Table 1 shows that the coefficient of variation (CV) for As is 1.3, suggesting a distribution that is not nearly-normal. This is confirmed by the Q–Q plot for As in Figure 6. Figures 1 and 2, and the histogram in Figure 5, show that the dataset is skewed to the right, implying that a Student's t-distribution is not appropriate for this dataset. A statistical software package confirmed this, and also determined that the Chebyshev inequality method produced an overly conservative UCL$\bar{x}$ for this dataset. The package recommended the use of a gamma distribution; this led to a calculated value for the 95% UCL$\bar{x}$ of 120.5 mg/kg.

The dataset for Zn has a CV of 2.2 and is highly skewed to the right, as can be seen from Figures 1 and 2, the Q–Q plot for Zn in Figure 11, and the histogram in Figure 5. The skewness suggests that a Student's t-distribution is not appropriate for this dataset. A statistical software package confirmed this, also finding that the dataset does not follow a discernible distribution. The package therefore recommended the use of the Chebyshev inequality method, which calculated a 95% UCL$\bar{x}$ of 214.7 mg/kg.

Appendix M: Worked example including results which are non-detects

Background

A site is proposed to be re-developed for recreational purposes as a park. An intrusive investigation has been carried out on a systematic square grid and all results were found to be less than the health investigation levels (HIL) and ecological investigation levels (EIL) for the proposed recreational use (HIL-C) except for cadmium which has a HIL-C criterion of 90 mg/kg. The underlying natural material was found to have no detections of cadmium greater than the limit of report (LOR). A QA/QC assessment has been performed and all data were found to be acceptable.

Results

The dataset shown below consists only of samples from the fill. Eight sample locations were assessed via testpits (testpit 1: TP01_0.2 and TP01_0.5, etc.). Samples were collected from different depths (reflected in the suffix of the sample identification so TP01_0.2 is the sample collected from 0.2 m below ground level at TP01). The field notes observed that the fill was highly heterogenous. The results are shown in Table 11.

Table 11 Results for cadmium

Sample ID	Cadmium (mg/kg)
TP01_0.2	<LOR
TP01_0.5	**95**
TP02_0.2	<LOR
TP02_0.5	<LOR
TP03_0.2	<LOR
TP03_0.5	**119**
TP04_0.2	<LOR
TP04_0.5	<LOR
TP05_0.2	<LOR
TP05_0.5	**93**
TP06_0.2	<LOR
TP06_0.5	<LOR
TP07_0.2	87
TP07_0.5	81
TP08_0.2	<LOR
TP08_0.05	<LOR

The data that is shown as <LOR is less than the limit of reporting (0.4 mg/kg).

There are 16 results in the dataset, of which 10 are less than the limit of reporting. Three results are greater than the assessment criteria. The consultant needs to decide if the material is suitable for the proposed site use or if they should recommend further sampling, a site-specific risk assessment, remediation or management.

Calculations and discussion

The NEPM (NEPC 2013) requires that the 95% UCL is less than the assessment criteria (90 mg/kg) AND the standard deviation is less than 50% of the assessment criteria (45 mg/kg) AND none are greater than 250% of the assessment criteria (225 mg/kg).

Initial calculations for 95% UCL and the standard deviation were performed using the 'substitution method' where zero, half the limit of reporting and the limit of reporting were each used, to substitute when the results were <LOR. The results are shown in Table 12.

Table 12 Calculations of 95% UCL and standard deviation for cadmium using the substitution method

Substitution value (mg/kg)	Percent of the limit of reporting	95%UCL (calculated using statistical software) (mg/kg)	Standard deviation (calculated using statistical software) (mg/kg)
0	0%	79.9	**46.1**
0.2	50%	**144**	**46.0**
0.4	100%	**144**	**45.9**

When considering the substituted results, there is a range of results that the consultant could report. The 0% substitution found that the 95% UCL was less than the criteria, but the standard deviation was greater than 50% of the assessment criterion. As well, the substitution with zero for the limit of reporting potentially underestimated the risk. The 95% UCL calculated for the 50% and the 100% substitutions exceeded the assessment criterion, as did the maximum result. The standard deviation for all substitutions was greater than 50% of the assessment criterion. After considering these results, the consultant might conclude that further sampling, a site-specific risk assessment, remediation or management under an EMP is required. A more sophisticated statistical treatment of the non-detects can be used to support decision making.

Using the same statistical software package (USEPA 2015a), the consultant entered the results so non-detects could be identified, rather than substitute a value of 0%, 50% or 100% of the limit of reporting. The software then calculated the 95% UCL and standard deviation, using a variety of statistical methods (USEPA 2006a) and recommended which values to use, based on considerations of the dataset's statistical characteristics. In this instance, values of 52 mg/kg and 44 mg/kg were obtained for the 95% UCL and the standard deviation, respectively.

This analysis was then used as one strand in a multiple line of evidence approach, where consideration was also given to the depth of the exceedances of the assessment criteria and the likelihood of receptors being exposed to the high cadmium-containing material. The consultant could rely on a more statistically defensible approach and conclude that the fill material was suitable to remain on-site without requiring remediation or management.