CEN BTTF 151

Date: 2007-04-03

CSS99058

prCEN/TR XXXX-1

Secretariat: DS

Sludge, treated biowaste, and soils in the landscape – Sampling – Part 1: Guidance on selection and application of criteria for sampling under various conditions

Einführendes Element — Haupt-Element — Ergänzendes Element

Élément introductif — Élément central — Élément complémentaire

ICS:

Descriptors:

Document type: Technical Report
Document subtype:
Document stage: CEN Enquiry
Document language: E

TR1_HZTL_Draft revised_03Apr07.doc

Contents

Page

Foreword

This Technical Report (prCEN/TR xxxx-1) has been prepared by Technical Committee CEN BT TF151 “Horizontal”, the secretariat of which is held by DS.

This document is currently submitted to the CEN Enquiry.

This document has been prepared under a mandate given to CEN by the European Commission and the European Free Trade Association, and supports essential requirements of EU Directive(s).

The following TCs have been involved in the preparation of this document:

CEN/TC292 Characterization of waste

This Technical Report is one of a series of five Technical Reports dealing with sampling techniques and procedures, and provides essential information and instructions for the application of the European Standard:

prEN xxxxx: Sampling of sludge, treated biowaste, and soils in the landscape – Sampling – Framework for the preparation and application of a sampling plan

The subject of the Framework Standard is the preparation of a sampling plan, within the framework of an overall testing programme as illustrated in Figure 1 of prEN xxxxx:date. The Framework Standard can be used to:

- produce standardized sampling plans for use in regular or routine circumstances;
- incorporate specific sampling requirements into national legislation;
- design and develop a sampling plan on a case by case basis.

The Technical Reports display a range of potential approaches and tools to enable the sampling plan to be tailored to a specific testing scenario. This approach allows flexibility in the selection of the sampling approach, sampling point, method of sampling and equipment used.

This Technical Report describes the statistical principles related to sampling, and provides methods based on these principles enabling a testing programme to be defined that will produce results sufficiently reliable for the decision-making process for which they are required.

Sludges, treated biowastes and soil arise in a wide variety of types (e.g. pastes, liquids, granular materials) and sampling situations (e.g. stockpiles, bags, fields). There can also be a variety of sampling objectives within each of three broad categories (basic characterisation, compliance testing and on-site verification). Consequently this Technical Report cannot provide definitive instructions for each and every case on the practical details of the testing programme, such as the required number of samples, the size of these samples, and whether they should be spot or composite samples. Instead, its aim is to expose the factors that influence the choice of these detailed components of the sampling exercise, and to provide statistical tools that can then be applied to determine the most appropriate testing programme for any given sampling scenario.

Introduction

Sludge and treated biowaste can be applied to land for the purpose of beneficial land use. The testing of sludge, treated biowaste and soil enables informed decisions to be made on whether land application is appropriate (or not). To undertake valid tests a (number of) representative sample(s) of the sludge, treated biowaste or land will be needed.

The subject of the Framework Standard prEN xxxxx is the preparation of a sampling plan, within the framework of an overall testing programme as illustrated in Figure 1 of prEN xxxxx:date.

The development of a sampling plan within this framework involves the progression through three steps or activities.

1) define the sampling plan;

2) take a field sample in accordance with the sampling plan;

3) transport the laboratory sample to the laboratory.

This Technical Report provides information to support Key Step 1 of the sampling plan process map and describes the selection of the sampling approach that can be used in the recovery of a sample for a wide variety of sludges, treated biowastes and soils in the landscape. Specifically this Technical Report provides information to support 4.2.8 (Select sampling approach) of the Framework Standard. Due consideration and selection of statistical criteria is of key importance in the production of a sampling plan as it provides the sole means of ensuring that, wherever possible, the type and number of samples taken will address a clearly identified objective and will provide results that achieve a tolerable level of reliability.

The main statistical steps that should be worked through in planning a sampling programme are given in Table 1. The information in Clauses 4 to 6 has been arranged so that it so that it follows the sequence of steps identified for definition of a sampling plan as discussed in the Framework Standard.

Table 1 – Main statistical steps in defining a sampling plan for a testing programme

Step	Subject
Specify the objective of the testing programme	
1	Specify the objective of the testing programme
Develop the technical goals from the objective	
2	Define the population to be sampled
3	Assess variability
4	Select the sampling approach
5	Identify the scale
6	Choose the required statistical approach
7	Choose the desired reliability
Determine the practical instructions	
8	Choose the sampling pattern
9	Determine the increment/ sample size
10	Determine the use of composite or individual samples
11	Determine required number of samples
Define the sampling plan	

12	Define the sampling plan

To illustrate the application of these principles, examples of sampling scenarios are provided in Annex E.

This Technical Report should be read in conjunction with the Framework Standard for the preparation and application of a sampling plan as well as the other Technical Reports that contain essential information to support the Framework Standard. The full series comprises:

prEN xxxxx: Sludge, treated biowaste, and soils in the landscape – Sampling – Framework for the preparation and application of a sampling plan

prEN ZZZZ: Sludge, treated biowaste, and soils in the landscape – Sampling – Vocabulary

prCEN/TR XXXX-1: Sludge, treated biowaste, and soils in the landscape – Sampling – Part 1: Guidance on selection and application of criteria for sampling under various conditions.

prCEN/TR XXXX-2: Sludge, treated biowaste, and soils in the landscape – Sampling – Part 2: Guidance on sampling techniques

prCEN/TR XXXX-3: Sludge, treated biowaste, and soils in the landscape – Sampling – Part 3: Guidance on sub-sampling in the field

prCEN/TR XXXX-4: Sludge, treated biowaste, and soils in the landscape – Sampling – Part 4: Guidance on procedures for sample packaging, storage, preservation, transport and delivery

prCEN/TR XXXX-5: Sludge, treated biowaste, and soils in the landscape – Sampling – Part 5: Guidance on the process of defining the sampling plan

The Technical Reports contain procedural options (as detailed in Figure 2 of prEN xxxxx:date) that can be selected to match the sampling requirements of any testing programme.

1 Scope

This Technical Report discusses the statistical principles of sampling, and provides a number of statistical tools to assist in the design of testing programmes for application to sampling under various conditions.

NOTE 1 Given the great variety of soil, sludge and treated biowaste types, sampling situations and objectives, this Technical Report cannot provide definitive instructions that cover all scenarios. Instead, it discusses the basic statistical approach to be followed, and provides statistical tools that can be applied to determine the amount and type of sampling (e.g. number of samples and sample size) in any given situation to achieve results of adequate reliability (i.e. precision and confidence).

NOTE 2 The document provides considerable detail on current best practice, but is not exhaustive.

NOTE 3 To reinforce the text, the document provides a number of worked examples.

2 Normative references

The following documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

prEN ZZZZ: Sludge, treated biowaste, and soils in the landscape – Sampling – Vocabulary

3 Terms and definitions

For the purposes of this document, the terms and definitions given in prEN ZZZZ apply.

NOTE B.1.1 provides an additional list of definitions that are specifically relevant to the various annexes.

4 Specify the objective of the testing programme

The objective of the testing programme should be an unambiguous statement of overall purpose. The objective should be made clear prior to selecting a sampling strategy, as it is an essential first step towards defining the type and quality of the information that is to be obtained through sampling. A clearly defined objective is required to identify the material population that will be characterized through sampling.

NOTE 1 A testing programme might have several objectives, but in most cases it will have only one. Each objective might result in a sampling programme; each sampling programme needing a separate sampling plan.

NOTE 2 Examples of possible objectives of the testing programme are to:

— compare the quality of the test material with quality levels defined in European or national legislation;

— characterize the test material following a change in ownership;

— determine the reusability of the test material;

— determine the leachability of the test material;

— assess the human health and / or environmental risks posed by the test material.

NOTE 3 Sampling will not be necessary in every case for meeting the objective. For example, the objective of an on-site verification might be simply to establish the identity of the material received. (Is it the correct material? Has it been delivered to the correct location for landspreading?)

In most cases, the objective is insufficiently detailed to produce the specific instructions needed in a sampling plan and needs to be translated into technical goals. These provide a more detailed specification of the sampling activity, and are sufficiently comprehensive to enable all aspects of the sampling plan to be

determined – the type, size, scale and number of samples to be taken, the way they are selected from the material under investigation, and so on. The process of developing the technical goals from the objective is discussed in detail in Clause 5.

5 Develop the technical goals from the objective

5.1 General

Once the objective(s) of the technical programme has been agreed, the next step is to develop the technical goals. This is a critical step because, once the technical goals have been defined, the specific sampling and data analysis requirements can be identified and the statistical analytical tools that will provide a consistent means of assessing and interpreting testing data can be identified. Such tools ultimately provide the means of verifying whether or not the technical goals have been met.

In some cases translating the objective into technical goals is straightforward because, for example, details such as the type of sample to be taken, or the statistical parameter to be determined from the results is specified in European or national legislation. Otherwise the technical goals will need to be defined in consultation with all involved parties. Conflicts can arise between (a) the desired reliability and scope of the sampling, and (b) the available resources; in this case compromise will be necessary. This makes it all the more essential that the involved parties agree on the technical goals and their implications prior to sampling..

Some technical goals might be sufficiently well-defined that they can be incorporated directly into the sampling plan (for example, the material to be sampled and the constituents to be tested). Other technical goals (for example the scale and confidence level) might require further ‘translation’ into practical instructions to the sampler.

The main steps in deriving the technical goals from the objectives are:

Define the population to be sampled See 5.2 and prCEN/TR xxxx-5

Assess variability See 5.3

Select the sampling approach See 5.4

Select constituents to be studied See prCEN/TR xxxx-5

Identify the scale See 5.5, Annex A and prCEN/TR xxxx-5

Choose the required statistical parameter See 5.6

Choose the desired reliability See 5.7

5.2 Define the population to be sampled

5.2.1 General

The term ‘population’ is a statistical term for defining the total volume of material about which information is required through sampling. Specification of the population should be one of the first steps in defining the sampling plan.

It is important to check in the process of defining the sampling plan that all involved parties are talking about ‘the same amount of material’.

5.2.2 Population

Commonly it is impractical to sample from the total population. It is customary, therefore, to define the ‘population’ as a convenient subset of the overall population that is believed to be typical of that wider overall

population. It is important to appreciate that an appropriate choice of population relies on *the experience and judgement of the interested parties*: it is not a statistical task.

It is also important to define the population explicitly over space and/or time; if this is not done, it is impossible to say whether a particular sampling exercise will result in representative samples.

NOTE For some sampling objectives, spatial variation might not be relevant (e.g. when sampling liquid from a pipeline at intervals through time), whilst for other objectives, temporal variation might not be relevant (e.g. when sampling from a number of heaps of dewatered sludge at a sewage treatment plant).

In defining the population for sampling it is important to consider the issue of 'scale' (see 5.5).

5.2.3 Sub-population

Cases arise where it is difficult or even impossible to sample certain parts of the population due to access restrictions. In such circumstances it is useful to define a subset of the population – termed the 'sub-population' – which restricts sampling to a more convenient region. The sub-population is therefore the specific part of the population that will be targeted for sampling, and which is thought to be sufficient to characterize the population. Sampling might therefore be carried out on either the population or sub-population depending on the volume of, and access to, the material under consideration.

The definition of a number of sub-populations might be useful where a large population is under investigation. These might be based on known changes in the production process or expected concentration levels. Alternatively the sub-population might be based on a characteristic of the material, such as any 'deviating parts' (e.g. white particles in a black material).

NOTE 1 A variety of terms could be used to define a sub-population according to the context, including 'lot', 'batch', 'big bags' or 'stockpile'. Whatever terms are used, their interpretation might be confusing and will be highly dependent on the definition of the testing programme. For stockpile sampling, for example, the population will often be the same as the lot or batch to be sampled, while a sub-population would be a part of that lot. In other cases a number of individual stockpiles might be related to each other – for example, consecutively produced batches of biowaste. The stockpiles might then be viewed collectively as the population, while an individual stockpile is the sub-population. Alternatively, it might be appropriate to define the collective of stockpiles as the overall population and each individual stockpile as a population.

NOTE 2 Given this risk of multiple interpretations, the Standard and TRs in this series only use the terms 'overall population', 'population' and 'sub-population'.

5.2.4 Examples

Some examples illustrating how it is possible to define overall population, population and sub-population for various categories of material are as follows:

EXAMPLE 1 LIQUIDS

Overall population: The total amount of liquid that passes through the slurry lagoon during a year.

Population: The total liquid held in a slurry lagoon on a particular date.

Sub-population: The volume of liquid accessible from a bridge across part of the lagoon.

EXAMPLE 2 SLUDGES

Overall population: The entire contents of all sludge tankers leaving a treatment works in a year.

Population: The entire contents of all sludge tankers leaving a treatment works in a particular week.

Sub-population: The columns of sludge accessible from the top inspection hatches of all tankers leaving the works in a particular week.

EXAMPLE 3 GRANULAR MATERIALS

Overall population: The entire production of treated biowaste at a composting plant during a year

Population: The contents of a compost heap over a specified area.

Sub-population: All material within 2 metres of the perimeter of the heap.

EXAMPLE 4 GRANULAR MATERIALS

Overall population: All thermally dried sludge granules produced at a sewage treatment works since it started operating.

Population: All thermally dried sludge granules produced at a sewage treatment works in a particular month

Sub-population: All thermally dried sludge granules produced at a sewage treatment works during the working day (e.g. 08:00 to 16:00).

EXAMPLE 5 SOILS

Overall population: All the land belonging to a farm.

Population: A field within the farm to which sludge is to be applied.

Sub-population: A part of the field that is less than 5 hectares in area and farmed for the same purpose.

It is important to appreciate that the resulting samples can only be representative in relation to the defined sub-population. Their relevance to the population is dependent on the validity of the assumptions made when defining the sub-population. Due consideration of scale is required when defining the sub-population (see 5.5 and especially Annex A).

The sampling plan should contain a specified description of the population or sub-population to be sampled to avoid possible ambiguities in sample collection.

Prior to sampling a check should be made that the material matches the detailed description of the population or sub-population as specified in the sampling plan.

NOTE Photographs of the material to be sampled might be taken to provide evidence of its identity if needed.

5.3 Assess variability

5.3.1 General

A key element in testing programme design is the need to understand the main components of variability in the population being sampled. In general, variability is a characteristic of the material that cannot be changed without intensive manipulation of the material, for example, stirring to mix a sludge that has settled into layers during storage. Its investigation is important because the more that is understood about the types of spatial and temporal variability affecting the material under investigation, the greater will be the opportunity for that knowledge to be exploited when designing the sampling programme.

EXAMPLE 1 Suppose a preliminary sampling exercise shows that day-to-day variation in the contents of bags of pelletized sludge is much greater than variation between the bags produced on any given day. If the objective of the testing programme is to characterize production over a longer period, the most reliable result would be obtained by taking a sample from a newly produced bag on as many days as possible in the period of interest rather than take samples from several bags produced on the same day.

EXAMPLE 2 Depending on the purpose of sampling, knowledge of a marked temporal cycle would give an option to (a) sample systematically over the cycle to smooth out that component of variation, or (b) target the sampling to the worst point in the cycle.

The impact of variability on the sampling exercise is heavily influenced by choice of scale of sampling – that is, the mass or volume of material that is taken into account to undertake an assessment of that material, where variations on a smaller scale than this are deemed to be unimportant. For more information see 5.5 and Annex A.

5.3.2 Spatial variability

5.3.2.1 General

Visualized in bulk, most materials exhibit some degree of heterogeneity. The origin of this spatial variability will often be the physically different locations from which the material has arisen. However, in other cases it might actually be due to *temporal* variations in the process producing the material. The spatial variability is an inherent characteristic of the population, which will not change without manipulation of the material (e.g. by stirring a bag of compost, or mixing a heap of dewatered sludge).

EXAMPLE 1 Spatial variation of metal concentrations within an agricultural field might reflect variations in the bedrock from which the soil has formed and/or additions from external sources, such as aerial deposition.

EXAMPLE 2 A major source of spatial variability within a heap of dewatered sludge might be the substantial differences from day to day in the average metal concentrations in the influent to the sewage treatment plant.

5.3.2.2 Within-stratum variability

Within-stratum variability is the term for the variation seen between samples taken from the same stratum – assuming that, in the case of sampling granular material, the sample size is sufficiently large for the effect of fundamental variability (see A.2) to be negligible. It is important to distinguish carefully between within-stratum and between-stratum variability, as their relative magnitudes have a critical bearing on how a given amount of sampling effort is best deployed.

5.3.2.3 Between-strata variability

Between-strata variability is the component of spatial variability that is introduced when there is spatial variability between different parts (strata) of the population.

NOTE 1 Examples of between-strata variation are the differences in the average concentration of a contaminant between:

- heaps of sludge in a holding area;
- heaps of compost produced from different source materials; and
- bags of material over a week's compost production.

NOTE 2 The distinction between within-stratum and between-strata variation is most obviously relevant when the material is in physically distinct parts. However, the concept of within-stratum variability is of equal relevance and importance to the testing programme design when the material arises or accumulates sequentially through time – as, for example, with material on a conveyor belt.

5.3.3 Temporal variability

5.3.3.1 General

Temporal variability can be considered as being of three main types: cyclic, driven, and random.

5.3.3.2 Cyclic variability

This is where the material characteristic exhibits a regular temporal pattern – for example, according to time of day, day of week, or time of year.

EXAMPLE In communities with large seasonal load changes, such as a holiday resort or where there is a food processing operation (fish, fruit, or vegetables), the characteristics of the sewage sludge might vary over the course of a year.

5.3.3.3 Driven variability

This is the term given to temporal variability that is caused or 'driven' by known factors.

EXAMPLE The characteristics of sewage sludge might fluctuate due to variation in rainfall patterns. During dry summer months solids settle in sewers and then get flushed through to the works in heavy rains. This changes the organic loadings on the sludge treatment works.

5.3.3.4 Random variability

Many other factors (mostly unknown) will additionally be influencing the material characteristics through time. Random variability is most often related to a (large) number of different (small) sources. The net effect of all these appears as unaccountable or random temporal variation.

NOTE Whenever temporal variability is expected, good knowledge is necessary of the production process of the material and its relation to the point and/or moment of sampling.

5.4 Select the sampling approach

5.4.1 General

There are two primary approaches to sampling. For the purposes of the Framework standard and this supporting TR these are termed 'probabilistic' and 'judgemental' sampling.

The use of probabilistic sampling should always take precedence where a quantifiable level of reliability is required in the results of the population being tested, because any deviation from probabilistic sampling will result in the loss of information on the reliability of the results.

5.4.2 Probabilistic sampling

The basis of probabilistic sampling is that each element within the population to be assessed has an equal chance of being selected by the sampling process. This implies that the whole population is accessible for sampling – even if, for example, it is a large stockpile. The key benefit of taking this 'statistical' approach to sampling is that the reliability of the resulting conclusions can be quantified (see 5.7.3).

The selection of appropriate sampling equipment is also important to ensure that a representative sample can be collected, for example, the entire particle size distribution for particulate materials.

NOTE Probabilistic sampling can be adopted in a stepwise approach: where the results from a sampling exercise are unacceptably imprecise, additional random samples can be taken to provide an improved measure of uncertainty. However, this approach will obviously increase the testing cost.

5.4.3 Judgemental sampling

With 'judgemental' sampling, in contrast to probabilistic sampling, samples are collected using at best a partially-probabilistic procedure, and at worst a non-probabilistic approach. The most common reason for selecting judgemental sampling is that representative sampling from the whole population is practically impossible, given the available resources in time and/or money. In addition, judgemental sampling might also be undertaken to deliberately target a specific item or point within the population (this type of sampling is commonly referred to as spot sampling).

The use of judgemental sampling will result in samples being taken from a sub-population, which is nearly always substantially more restrictive than the whole population. Within that sub-population, however, it might be feasible for the sampling to be probabilistic. This option should be adopted wherever possible, as it will mean that the results are at least representative for the part of the population sampled – though they still of course run the risk of being biased for the whole population.

NOTE 1 For example, samples might be taken at random from the top 50 cm of a stockpile. The advantage of doing this is that it allows statistically sound information to be generated for at least the sub-population sampled. This makes it easier to assess the possible errors involved in extrapolating to the whole population (i.e. stockpile), whilst also making explicit the way in which the sampling is unrepresentative. Errors should also be assessed in the light of available knowledge for the methodology adopted.

NOTE 2 The adoption of judgemental sampling at this level might therefore have severe financial and/or environmental consequences.

Given these unquantifiable uncertainties, the usefulness of the results from judgemental sampling is highly dependent on the reliability of the background information – on which any expert judgement, and ultimately the sampling plan, is based. The limitations of judgemental sampling will therefore be especially acute for new sampling scenarios where there is an absence of relevant information or validation results.

5.5 Identify the scale

The 'scale' is a crucially important element in defining a sampling programme. It defines the minimum quantity (mass or volume) of material below which variations are judged to be unimportant. For example, if the scale of sampling is defined to be 'a bag of composted biowaste', then variations in any characteristic of the compost within the volume of a bag are declared to be of no concern. The scale of sampling is discussed more fully in Annex A.

The amount of spatial variability in the population cannot be quantified without defining the scale on which that variability occurs. For example, the variability from gram to gram of material in a sub-population is likely to be larger than the variability from kilogram to kilogram. If variations in concentration on so fine a scale as this are believed to be important, then that is the scale on which the sampling should operate. If, conversely, concentration variations within any one kilogram of material are irrelevant, the primary aim of the sampling should be to quantify variability solely on the kilogram-to-kilogram scale. It is therefore of vital importance that the scale is stated explicitly.

NOTE When the scale of interest is 1 kg, a much higher degree of variability might be expected within the stockpile than for a scale of, for example, 5 tons. Thus, if the purpose of sampling were to test compliance with a limit, the resulting data might lead to rejection of the stockpile on the scale of 1 kg, but acceptance of the stockpile on the scale of 5 tons.

Conversely, a given scale represents the quantity or magnitude of material on which you intend to base your measurement or that the measurement needs to relate to.

It follows that when obtaining information about a material at the specified scale, each numerical value is a mean for the volume or mass of material at that scale.

5.6 Choose the required statistical parameters

A 'statistical parameter' is any numerical characteristic of a population – for example, its mean or its standard deviation. A key step in planning a testing programme is to specify the statistical parameters that are required to be estimated. It is important to do this because the choice generally has a critical bearing on both the type of sampling and the number of samples needed.

For a number of commonly used parameters, Annex B provides statistical equations both for estimating the parameter itself, and for calculating the uncertainty associated with that estimate. The second of these is a critical piece of information, because it provides the quantitative link between the number of samples and the achievable reliability (see 5.7 and Annex C).

For estimating percentiles, as Annex B indicates, the choice of method depends on what can be assumed about the underlying 'probability distribution' – a statistical term used to describe the relative frequencies with which different values arise in a given population. Two probability distributions are particularly useful – the normal and lognormal distributions. Annex B provides a brief description of these. It also introduces the binomial distribution because of its importance to the handling and interpretation of 'presence/absence' data.

The objective of the testing programme will guide the selection of the most appropriate statistical parameter. Three generic levels of testing are commonly distinguished:

— basic characterization;

— compliance testing;

— on-site verification.

NOTE 1 Other definitions might apply.

When testing is aimed at basic characterization, the investigation is likely to require measures of (a) variability and (b) extreme behaviour of key constituents. A large number of samples might be needed to meet these requirements. However, such a sampling exercise would also be useful more widely in providing a good indication of the overall statistical distribution of those key constituents.

Depending on the statistical distribution the mean might not provide the most useful estimate of the 'central characteristic'. For example, where the statistical distribution is positively skewed, the median (or 50-percentile) might be a more useful estimator than the mean. This will depend on the objective of the testing programme.

Where it is more appropriate to estimate the extreme values of a statistical distribution, a percentile might be the preferred choice of parameter (see Annex B).

When the objective is compliance testing, the choice of statistical parameter will usually have already been designated by the compliance rules defined by the regulator, as will the scale at which the material is required to comply (see Annex A). Commonly mean and percentile values are used in this type of testing.

In the case of on-site verification the investigation will focus on either the measured value in relation to a compliance level that should not be exceeded (making it a simple type of compliance testing), or simple 'presence/absence' attribute measurements.

NOTE 2 On-site verification does not necessarily require actual (analytical) measurements. It might well need only a visual inspection of the material in order to determine if it is indeed the type of material that was expected.

5.7 Choose the desired reliability

5.7.1 General

The reliability of a testing programme is a general term embracing three statistical concepts: 'bias', 'precision', and 'confidence'. The objective of the testing programme will influence the degree of reliability that is regarded as acceptable, but the final selection of reliability criteria will nearly always need to be a compromise between cost and expectation. The process of defining the sampling plan therefore, might well be an iterative process.

Given the important decisions that are likely to rest on the findings of a basic characterization exercise, it is suggested that the reliability should be as high as possible. Conversely, given the 'quick check' format envisaged for on-site verification, the achievable reliability for any one assessment will in many cases be low. However, this could be offset to some extent where a large number of similar checks are available.

5.7.2 Precision and Confidence

A unique property of probabilistic sampling (see 5.4) is that it allows an error band – known as a 'confidence interval' – to be placed around any parameter estimate. The semi-width of the confidence interval is usually known as the 'precision'. This depends on:

— the desired degree of confidence;

— the variability in the population or sub-population;

— the sampling pattern (see 6.2);

— the chosen number of samples; and

— the assumed statistical probability distribution followed by the population (see Annex A).

The key benefit of being able to estimate the achievable confidence and precision associated with any proposed testing programme is that it provides a quantitative link between the sampling resources used and the reliability of the resulting answers.

5.7.3 Errors in the testing programme

— **Systematic error** (also known as bias). A testing programme with a systematic error is one that has a persistent tendency either to under-estimate or to over-estimate the parameter of interest. Systematic errors might easily occur when sampling takes place from a sub-population.

— **Random error.** As the testing programme never samples more than a very small fraction of the whole population, the composition of the samples is to some degree determined by chance. Consequently the composition of the sample will never be exactly the same as the composition of the whole population. The difference between sample and population resulting from this chance process is known as the random error.

— **Statistical sampling error.** Statistical sampling error, or more commonly 'sampling error', occurs as a consequence of the fact that only part of the population is sampled. Consequently the calculated characteristic will differ from the 'true' value of the whole population – that is, the value that would have been obtained if the whole population could have been sampled. This difference is known as the sampling error. This might either be a systematic error or a random error (or a combination of both), depending on the adopted sampling procedure. For a correctly applied probabilistic sampling exercise, the sampling error will be due solely to random error.

— **Physical sampling error.** In addition to statistical sampling error, which arises as an inevitable consequence of the random sample selection process, the sampling activity might itself introduce an additional error. This can be termed 'physical sampling error', and might take the form of either systematic or random error (or a combination of both). To minimize this type of sampling error, the most appropriate sampling device for the task should be selected, and the standards laid down for its correct use adhered to.

EXAMPLE 1 Suppose a cross-sectional sample is taken from the entire width of a conveyor, and the scoop fails to catch all of the fines (perhaps because of unevenness of the surface). The resulting sample will tend to consistently under-represent the fines component of the material (systematic error).

EXAMPLE 2 Suppose a mixed sample is taken from the end of a pipeline by collecting portions of the exit stream at regular intervals. This technique assumes that flow is constant over the whole time of sampling; thus error will be introduced if flow is in fact varying over the period. If the variations in flow are random, then the physical sampling error will be random in nature, but in most situations the resulting errors are likely to be both random and systematic.

EXAMPLE 3 Suppose a 3 cm diameter auger is used when sampling a heap of composted biowaste in which particle size might be as large as 5 cm. The larger particles cannot be sampled and will therefore have no contribution to the measurements. As a consequence there will be a systematic error in the measurements.

— **Analytical error.** Analytical error is the collective term for the errors that arise during the analytical activities necessary to obtain the desired results, including the sample pre-treatment, extraction or destruction of the sample and the subsequent analysis of the extract, destruate or eluate. A reliable estimate of the random component of analytical error, and an upper limit on the possible bias, will generally be available from the laboratory through its Analytical Quality Control (AQC) procedures.

6 Determine the Practical Instructions

6.1 General

Clause 5 has discussed the steps needed to develop the objective of the sampling plan into a number of more detailed technical goals. The technical goals should now be translated into practical instructions that are given to the sampler prior to sampling. The practical issues that should be considered in identifying these instructions are as follows:

— choose the sampling pattern See 6.2

— determine the minimum increment and sample size See 6.3

— the use of composite versus individual samples See 6.4

— determine the required number of increments and samples See 6.5

6.2 Sampling pattern

6.2.1 General

The sampling pattern defines where, when and how the required samples are selected from the population. Three probabilistic sampling patterns and two options for judgemental sampling are illustrated in Figure 1.

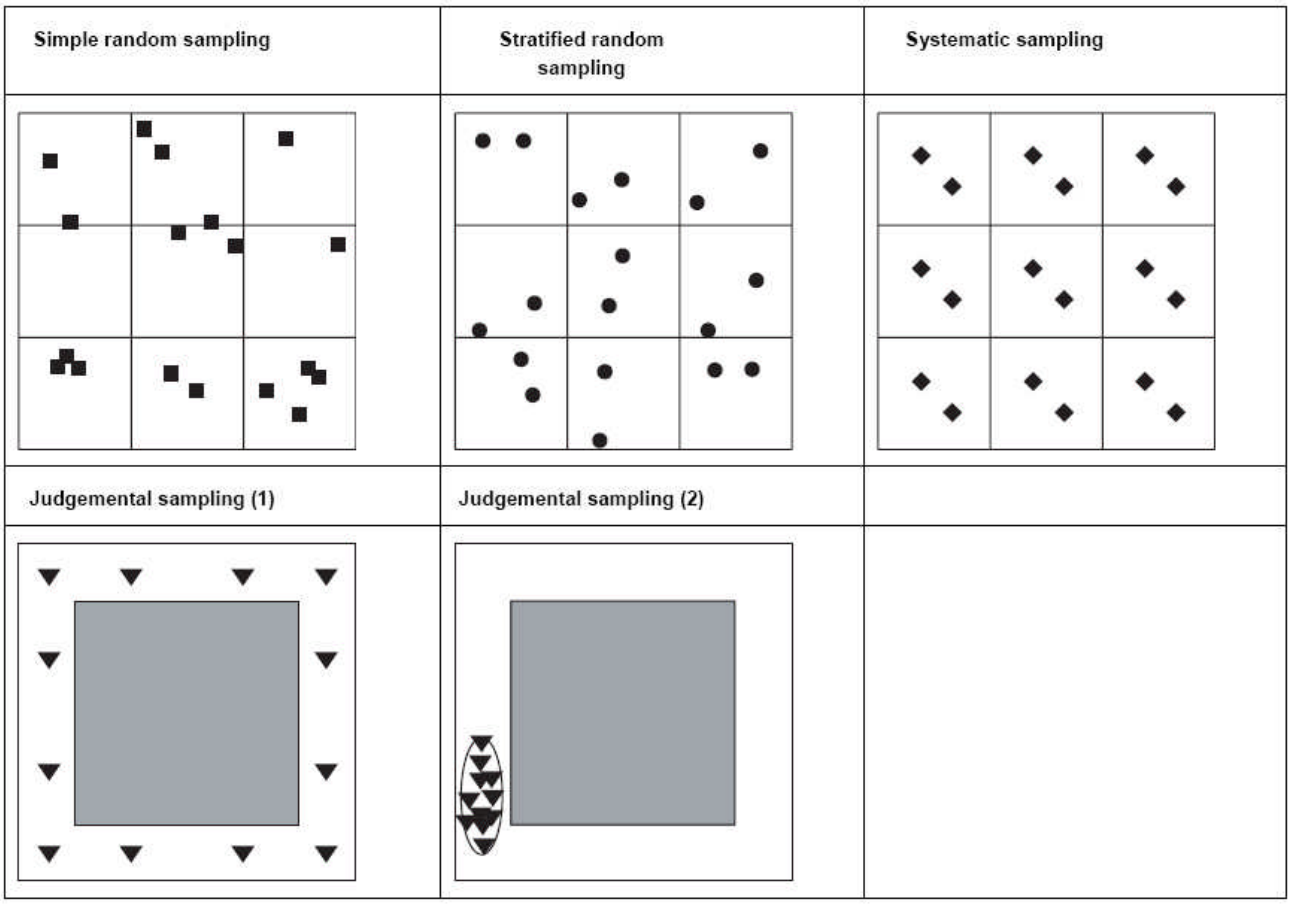


Figure 1 — Possible patterns of sampling

NOTE The figure illustrates the patterns for the context of a two-dimensional spatial area. However, the concepts apply as equally to temporal as they do to spatial components of variability.

6.2.2 Simple random sampling

With simple random sampling, every portion of the population has the same (small) chance of being selected as a sample. However, the resulting samples will not necessarily be very evenly spread across the population. Consequently other more structured forms of sampling are often preferred to simple random sampling.

6.2.3 Stratified random sampling

With stratified random sampling, specified numbers of samples are spread randomly over each of a number of strata that are predefined in the population. This preserves the advantages of random sampling (that is, every portion of the population has a known chance of being selected as a sample), whilst ensuring that each stratum is represented by a predetermined number of samples. Where the number of samples in each stratum is proportional to the proportion of the population falling into that stratum, the sampling is termed ‘self-weighting’. Often, however, there are advantages in having equal numbers of samples in each stratum, and subsequently weighting the results by the estimated stratum sizes in the population.

EXAMPLE Suppose a tank contains a liquid that has stratified into three layers: top (20%), middle (70%) and bottom (10%). Three samples taken from each layer give rise to mean concentrations of 8, 10 and 25 mg/kg. The overall mean concentration would be estimated by (20×8 + 70×10 + 10×25) /100 = 11,1 mg/l. Note that the reliability of the estimated mean concentration very much depends on the accuracy of the estimated stratification.

6.2.4 Systematic sampling

With systematic sampling the samples are evenly spaced across the population, starting from a randomly chosen point (to ensure that each item in the population has an equal chance of being sampled, to fulfil the requirements of probabilistic sampling). This has obvious operational advantages. For the benefits of probabilistic sampling still to apply, however, the approach does rely on the assumption that there are no systematic components of variation within the population that 'run in step with' the chosen sampling frequency.

EXAMPLE 1 Suppose a sample is taken from a production process on the second Tuesday in every month. This would generate a systematic error in the results if the process happened to follow a regular weekly cycle. This could be avoided, however, by sampling on the second day in each month.

EXAMPLE 2 Suppose the plan were to sample every 20th bag of material from a conveyor belt. This would produce a systematic error in the results if the process generated a higher metal concentration at a regular rate of one every five bags. The higher concentration would either be permanently missed or be over-represented in the resulting samples, according to when the systematic programme happened to start.

Systematic sampling should therefore be applied with care when it is used in place of a random or stratified random sampling.

6.2.5 Judgemental sampling

Judgemental sampling can embrace a wide variety of sampling patterns that essentially differ in terms of how far they deviate from a truly probabilistic approach.

Option (1) in Figure 1 shows a form of judgemental sampling that is based on a probabilistic approach for part of the population. The sampled sub-population is the narrow strip around the shaded region. Within this, however, there is a systematic sampling pattern (chosen such that there is no risk of the samples running in step with any systematic pattern that might be present within the sub-population). As this is a form of probabilistic sampling, the statistical benefits associated with this approach might be exploited. That is, the methodology of Annex B can be used both to estimate the parameter of interest and also to calculate a confidence interval to quantify the uncertainty surrounding that estimate. Of course these calculations are only valid for the sampled sub-population; the wider relevance of the sampling results depends entirely on whether or not this sub-population is representative of the whole population.

In contrast, the pattern in option (2) is fully judgemental. It provides no information about the waste except in the immediate vicinity of the samples, and so nothing can reliably be inferred about the quality of the sub-population. Conversely in some situations judgemental sampling can be the most appropriate form of sampling. For example, when the purpose of sampling is simply to investigate and estimate the characteristics of an atypical material that is unexpectedly present in the population.

6.3 Determine the increment and sample size (mass / volume)

6.3.1 General

An increment is the amount of material (mass or volume) that is obtained through one single sampling action. It is not analysed as an individual unit, but is combined with other increments to form a composite sample. Conversely a sample is an increment that does get analysed individually.

The degree to which account needs to be taken of the increment and sample size will very much depend on the type of material sampled. The minimum increment size is governed by the need for the sampling device to accommodate all particle sizes. Thus it has particular consequences for the sampling of particulate materials. In contrast, there is no practical requirement for a minimum increment size in the sampling of liquids, where the particle size goes down to the molecular scale.

A sample should be sufficiently large in order to minimize or exclude errors caused by the fundamental variability of the material that is determined by the differences between individual particles.

The terms fundamental variability (see A.2) and heterogeneity due to 'clustering' should not be confused. The latter relates to the preferential presence of a specific type of material to be in a specific part of the population, and can be dealt with by the sampling pattern (see 6.2). Fundamental variability, however, should be overcome by putting a demand on the sample size and hence the number of particles in a sample.

The following clauses provide information on the determination of minimum increment and sample size for a range of material types.

6.3.2 Liquids

As previously stated, the minimum increment and sample size have no specific relevance to the sampling plan design for liquids as the potential differences are at a molecular scale when compared to the size of the samples. When taking composite samples, the sample size will be governed by the number of increments and the increment size. The increment itself will be determined by the dimensions of the sampling equipment.

NOTE Often some sludges and treated biowastes (digestates) are considered to be liquids. However, these materials are in fact suspensions (a mixture of fluid and particles) and rarely can be considered fully mixed. This has consequences for sampling, particularly when the materials are stored and the particles have time to settle into layers.

6.3.3 Powders and sludges

Powders and sludges are basically particulate materials with a (very) small particle size; sludges also contain a substantial amount of liquid. Provided the sample device allows the entry of all particles present in the material being sampled there are no additional requirements for the minimum increment size.

Similarly, given the small size of the particles in these types of material the differences between individual particles will not have a major affect on the characteristics of a sample, as in practice, the sample will be large enough to consist of a (very) high number of particles. There are therefore no practical requirements for the minimum sample size.

The sample size will therefore be governed by the quantity of material required by the laboratory for analysis, whilst the dimensions of the sampling device will determine the increment size. As with liquids, the size of any composite samples will primarily depend on the number of increments and the increment size.

NOTE Although the distinction between a powder and a granular material is not always obvious, the consequences on the minimum increment and sample size are potentially great. Care should be taken, therefore, that the aperture of the sampling device is suitable for the particle size distribution in the material to be sampled.

6.3.4 Particulate / granular materials

Where samples are to be taken from a particulate or granular material, account should be taken of the *minimum* increment size. The size of the opening of the sampling device should be large enough to allow the entry of all particles present in the material. The aperture of the sampling device should also be large enough to allow the simultaneous entry of all particles within the material. In practice this means that the device opening should be at least three times the diameter of the largest particles. For a three dimensional sampling device the volume of the increment should be equal to $(3d)^3 = 27d^3$. For practical reasons, the diameter of the largest particles can be substituted by the size of the (estimated) 95-percentile of the particle size distribution.

More details on the minimum increment size are provided in D.2.2.

For particulate or granular materials the composition of individual particles could have a substantial influence on the composition of the sample, and the minimum sample size should be large enough to compensate for this. This is particularly important when the contaminant or characteristic of interest constitutes only a small proportion of the material. D.2.3. provides an equation for calculating the minimum sample size (by mass).

NOTE It is assumed that, in characterizing the material, we are interested not in the composition of the individual particles, but in the average composition of the material (at the specified scale). To measure that average composition the sample should contain a sufficient number of particles to ensure that the effect of any individual particle within the sample does not have a disproportionate effect on the total composition of the sample.

The actual size of the increments and samples will depend not only on the minimum increment and sample size but also on:

— the quantity of material required by the laboratory for analysis;

— the number of increments in a composite sample (when increments are taken); and

— the relation between the mass of the minimum increment size and the minimum sample size (in relation to the number of increments in a composite sample).

6.4 The use of composite versus individual samples

The objective of the testing programme, and in particular the choice of statistical parameter, will dictate whether individual or composite samples will generate the more appropriate type of data.

NOTE 1 There is an important distinction between an *increment* – which forms a part of a *composite sample* – and a *sample*, which is produced by a single sampling operation.

A number of basic scenarios are envisaged:

— When an approximate indication of the quality of a material is sufficient to meet the testing objective, as for example with on-site verification, this could be satisfied by the collection of one sample or at the most a small number of *samples*. In this example the costs of sampling and analysis would be low.

— Conversely a substantial number of *increments* should be taken if a reliable estimate of *mean* quality is required for one or a number of composite samples. Such an approach might provide a satisfactory approach for compliance testing (where the compliance value is relevant to a mean concentration). In this example the cost of sampling could be relatively high, but costs of analysis would be low.

NOTE 2 Although the use of composite samples based on a (relatively) large number of increments is an attractive option for obtaining a good estimate of the mean concentration without substantial analysis costs, chemical or physical restrains to adding increments might need to be taken into account, as for example where the quantification of volatile components is of importance.

— When a substantial number of *samples* are taken in order to obtain a reliable estimate of a specific quality of the material and also provide information on the degree of heterogeneity within the material – as for example in basic characterization – the costs of both sampling and analyses will be high.

6.5 Determine the required number of increments and samples

The required numbers of increments and samples should be determined using the methodology set out in Annex C.

7 Define the sampling plan

The previous clauses of this Technical Report all contribute to the definition of the sampling plan as specified in prEN xxxxx. The various choices and decisions that have been made using the principles outlined in these clauses can now be drawn together to complete the 'sampling methodology' section of the sampling plan as detailed in the example in Table A.1, Annex A of the Framework Standard, thereby developing a situation-specific sampling plan. In many cases an iterative process will be needed in reaching the finally agreed version of the sampling plan. This will ensure that a satisfactory compromise is reached between the objective as originally desired, and the objective that is practically achievable for the available resources in the light of any practical constraints of access and sampling.

The process steps that contribute to the definition of the sampling plan as specified in prEN xxxxx are:

Specify the objective of the testing programme (4)

- specify the objective of the testing programme (4)

Develop the technical goals from the objective (5)

- define the population to be sampled (5.2)
- assess variability (5.3)
- select the sampling approach (5.4)
- identify the scale (5.5)
- choose the required statistical approach (5.6)
- choose the desired reliability (5.7)

Determine the practical instructions (6)

- choose the sampling pattern (6.2)
- determine the increment/ sample size (6.3)
- determine the use of composite or individual samples (6.4)
- determine required number of samples (6.5)

Define the sampling plan (7)

- define the sampling plan (7)

NOTE Annex E provides a number of illustrative examples laid out in a standard tabular format.

Annex A
(informative)
The scale of sampling

A.1 Scale

Scale is one of the essential issues of sampling. The scale defines the volume or mass of material that a sample directly represents. This implies that when the assessment of the material is needed for example on one cubic metre, the sampling results should provide information on a cubic metre scale. Thus the analytical results should be representative for a cubic metre of material.

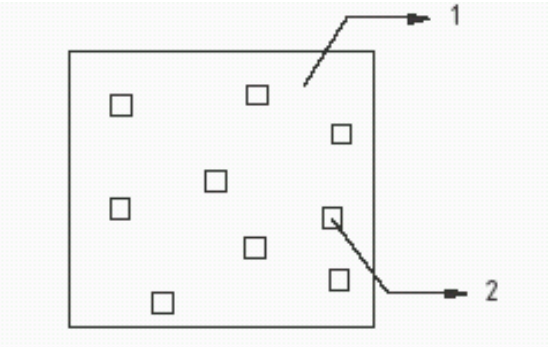
Depending on the objective of the testing programme, the scale of sampling might be equal to the size of individual particles of the material (for particulate materials), the size of the sub-population or even the whole population.

Scale can also be defined in terms of time: if the population is the total amount of material produced in one year, the scale might be one year (the whole population) but also one month, week or day, depending on the objective of the testing programme.

Defining the scale is important, as heterogeneity is a scale dependent characteristic. Assume a particulate material that consists of small particles that only vary in colour. The particles in the material are fully mixed. In a series of samples, each with the size of an individual particle, each sample will have a different colour. Therefore the observed heterogeneity in colour between these samples will be high. However, the degree of heterogeneity on a scale of for example, 1 kilogram, consisting of several thousands of particles, will be low. Each of these samples will have approximately the same mix of colours, and – looking from some distance (thus really on the scale of 1 kilogram) – the samples will have the same mixed colour. Thus the observed heterogeneity will now be low.

As a consequence of the direct relation between scale and heterogeneity, sampling results are only valid for the scale that is equal to the scale of sampling or higher scales. In general, the degree of heterogeneity will be higher for a smaller scale of sampling and will be lower for a larger scale of sampling.

Three specific examples for which the scale is defined are as follows:



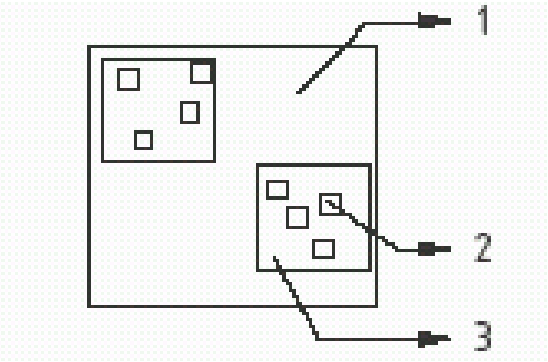
Key
1 Population 2.000 ton
2 Increment 200 gram (50 increments in a composite sample of 10 kg)

Figure A.1 – Scale situation 1

Situation 1 describes a population of 2.000 tons from which 50 increments are randomly taken. The resulting composite sample is 10 kg.

Assuming that the composite sample resulting from these 50 increments represents a good estimate of the mean concentration (but not of the variability) of the whole population, **the scale for the composite sample** in this example **is 2.000 tons**.

Note that the variability of the population (on the scale of the increments) is fully incorporated in the composite sample; the sampling method will however provide no information on the variability.

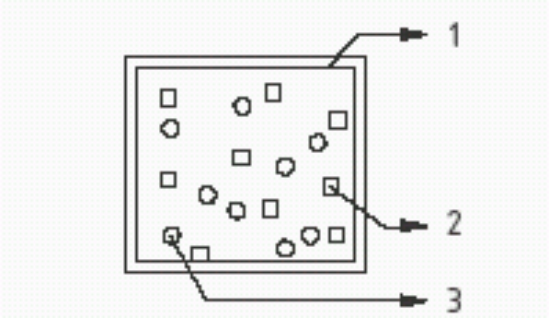


Key
1 Population 2.000 ton
2 Increment 200 gram (50 increments in a composite sample of 10 kg)
3 Sub-population 50 ton

Figure A.2 – Scale situation 2

Situation 2 describes a population of 2.000 tons. Within this population – perhaps only for the purpose of sampling – sub-populations are defined of 50 tons each. From each sub-population 50 increments are taken. The resulting composite samples are 10 kg, each representing a sub-population.

The mass represented by each composite sample is now the mass of the individual sub-populations; thus 50 tons. **The scale for each composite sample** in this example **is 50 tons**. The mean value of all composite samples yields an estimate of the mean concentration of the whole population of 2.000 tons and the variability within the whole population is estimated on a scale of 50 tons.



Key
1 Population 2.000 ton
2 Increment 200 gram (50 increments in a composite sample of 10 kg)
3 Increment 200 gram (50 increments in a composite sample of 10 kg)

Figure A.3 – Scale situation 3

Situation 3 describes a population of 2.000 tons. More than one composite sample is taken. However, each composite sample (existing of 50 increments) is obtained by taking random increments throughout the whole population. The mass represented by each composite sample is now equal to the mass of the whole population; thus 2.000 tons.

The scale for each composite sample in this example **is 2.000 tons**. The mean value of all composite samples yields an estimate of the mean concentration and the variability of the whole population of 2.000 tons is estimated on a scale of 200 grams (the mass of the increments).

The following example illustrates the effects of different definitions of the scale of sampling. Depending on the objective of the testing programme, the involved parties need to make a choice.

Consider the three sub-populations as shown in Table A.1. Each sub-population consists of thirteen individual parts that have a 'quality' that is symbolized by a number between 0 and 99. Heterogeneity is quantified by the coefficient of variation: a high coefficient of variation indicates a high heterogeneity.

When the scale of sampling is equal to the size of the sub-population, the sampling result will only be an estimate of the mean concentration for each sub-population. Comparing the sub-populations shown in Table A.1, sub-population 1 and 2 are comparable while sub-population 3 has a higher mean.

When the scale of sampling is equal to the individual parts within each sub-population, an estimate for the mean concentration of the sub-population is obtained in addition to an estimate for the heterogeneity within that sub-population. Comparing the sub-populations shown in Table A.1, now still gives the same result for the mean of the whole sub-population, but additionally it is shown that sub-population 2 has a higher degree of variability than sub-populations 1 and 3.

Table A.1— Example of three different sub-populations, characterized on the individual samples, the mean and coefficient of variation (CV)

	Subpop. 1	**Subpop. 2**	**Subpop. 3**	
	20	15	32	
	30	14	36	
	20	22	3	
	30	72	37	
	40	9	38	
	20	23	36	
	30	64	37	
	30	46	30	
	40	5	40	
	20	16	41	
	10	2	17	
	20	17	39	
	30	35	36	
				Population
Mean	26,2	26,2	32,5	28,3
Coefficient of variation	33,3%	84,2%	33,2%	

NOTE A high CV indicates a heterogeneous sample.

Finally, when the scale of sampling is equal to the total population we obtain only an estimate of the mean for the whole population.

Different choices can now be made on the scale of sampling:

The scale of sampling is equal to the scale of the individual parts. It is not possible to define a smaller scale of sampling. The result of this definition of the scale is that information on the heterogeneity within the sub-populations can be obtained by calculating (for example) the coefficient of variation. Additionally, the

heterogeneity between the sub-populations and within the population can be calculated. In this approach, the presumptions that led to identification of the sub-population as a relatively homogeneous part of the population can be verified. For example, it might be argued that sub-population 2 in Table A.1 is so heterogeneous that at least a part of sub-population 2 will not comply with certain quality standards, although the mean value is within the quality range. Many sub-populations of high heterogeneity might lead to a re-evaluation of the sampling plan. An important disadvantage is the cost for measuring the individual parts, in this case thirteen per sub-population[1].

The scale of sampling is equal to the scale of the sub-populations. Therefore no information on individual parts within a sub-population is gathered. Characterization of the sub-population is done by means of a composite sample per sub-population in which more than one of the individual items is put together prior to analysis. If this composite sample is taken and analysed correctly, the result of the composite sample will be a good estimate of the true mean of the sub-population. An important advantage of this approach is the low cost for measuring. An important disadvantage is the assumption that a composite sample can be obtained without a considerable sampling error. The analysis of a composite sample might pose problems as the amount of material in the sample will be (much) larger than the amount of material needed for the analysis and thus proper sample pre-treatment is necessary to obtain a representative analytical sample from a – potentially – highly heterogeneous composite sample. Additionally, there will be no information available on the heterogeneity within a sub-population.

The scale of sampling is equal to the scale of the population. In the example shown in Table A.1 the population is defined as the combination of the three sub-populations. Individual parts are gathered from the involved sub-populations and put together in a composite sample. Now there will be no information available on a smaller scale than the scale of the population. An important advantage is the (very) low cost for measuring, while, as long as it is technically possible to mix a large number of these parts, the result of the composite sample will still be representative for the true mean of the total population. But the population has to be treated as one entity. In case of a heterogeneous population (for example sub-population 2 in Table A.1) sampling on the scale of sub-populations or individual parts would have given the involved parties information that might have led to different choices for the destination of sub-populations of different quality.

Given the relation between scale and the encountered degree of heterogeneity, the applied scale of sampling might determine if a material is considered homogeneous (i.e. there is little variation between individual sample results) or heterogeneous (i.e. high variation between sample results).

The type of information that is desired, the possible destination, the financial means available and the technical possibilities of working with composite samples determine the choice on the scale of sampling.

In addition to the more technical perspective from which the definition of scale was described in the previous text, the scale of sampling can also (or even should) be defined by policy considerations. In principle the scale of sampling should be equal to the amount of material that is considered relevant from a policy perspective. An example of a policy-defined scale of sampling might be as follows:

Based on the radius of action of small animals living in soil, the mean concentration of a soil volume of 25 m^3 is considered as relevant for assessing the seriousness of soil contamination. It is assumed that these animals throughout their whole life span are exposed to the mean concentration of the pollutants in this soil volume. Thus, when assessing the seriousness of polluted soil, we are interested in the mean concentration within this volume of 25 m^3. When acute exposure to (very) high concentrations is considered not to be relevant, there is no need to gather information on a smaller scale than 25 m^3. The scale of sampling is therefore 25 m^3 and is achieved by taking a number of increments within this volume; an estimate of the true mean concentration on the scale of 25 m^3 can thus be obtained.

For the definition of scale in time one should consider a production process which results in a continuous stream of dewatered sludge. At t = 0 the production starts. Sampling takes place between t = 20 and t = 30. The sampling process results in a (good) estimate of the mean concentration between t = 20 and t = 30. Therefore, the scale of the obtained result is 10 (seconds, hours, …). Of course, knowing the mass of material

---

1 Note that it is not necessary (nor practical) to measure each individual item within a sub-population. A sample survey within each sub-population might be sufficient.

produced in for example 10 minutes, this time defined scale can be easily transferred into a mass defined scale.

A.2 Fundamental variability

Granular material will generally consist of different types and shapes of particles. As a consequence there is a degree of variability on the scale of the individual particles. This variability cannot be reduced without particle size reduction. This is called the 'fundamental variability'. It will be the cause of variability between samples whenever the characteristic of interest – e.g. the concentration of metals, or organic matter – is directly related to a specific portion or subset of the particles. Also when the concentration of the constituent of interest varies over the different particles, there is fundamental variability.

As the average number of particles per sample increases, so the effect of the fundamental variability becomes less dominant. Nevertheless, the effect can remain large even with a large number of particles in the sample if the constituent of interest (e.g. copper occurring incidentally within a material) arises in only a small proportion of particles but at very high concentrations. Annex D provides the details of a method that can be used to estimate the minimum size of samples to ensure that the error due to fundamental variability is as small as required.

Theoretically, fundamental variability on the scale of the individual particle also applies to liquids in which different substances are dissolved. However, as the particle size is on the molecular scale, the sample will always be large enough to contain a (very) large number of particles (molecules) and therefore fundamental variability is of no interest for liquids.

Annex B
(informative)
Statistical methods for characterising a population

B.1 Terms, definitions and symbols

B.1.1 Terms and definitions

B.1.1.1
binomial distribution
type of probability distribution that describes the statistical behaviour of 'presence/absence' data

NOTE If the presence or absence of some attribute (e.g. a limit exceeded) is noted for each of n random samples, and that attribute has an underlying probability of occurrence p, then the binomial distribution B(n,p) describes the variability to be expected in the observed number of samples showing the attribute.

B.1.1.2
composite sampling
process of taking composite samples

B.1.1.3
confidence level
value 100(1 - α) of the percentage probability associated with a confidence interval

NOTE Where α is the significance level.

B.1.1.4
confidence limits
each of the limits, T1 *and* T2, of the two-sided confidence interval or the limit T of the one-sided confidence interval

B.1.1.5
histogram
graphical representation of the frequency distribution consisting of a set of contiguous rectangles, each with a base width equal to the class width and area proportional to the class frequency

[ISO 3534-1:2006, definition 1.61]

B.1.1.6
lognormal distribution
family of probability distributions characterized by right-handed skewness, and often a useful approximation to environmental variability. So called because the *logarithm* of such a constituent is *normally* distributed

B.1.1.7
non-parametric method (distribution-free method)
any statistical method which makes no assumption about the probability distribution describing the variability of the sampled population, but is instead based on properties of the *ranked order* of the data

EXAMPLE For example, given seven sample values {22, 28, 29, 33, 37, 41, 66}, the median (50%ile) can be estimated by the 4th ranked value, namely 33.

B.1.1.8
normal distribution
continuous distribution having the probability density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}$$

where $-\infty < x < +\infty$ and with parameters $-\infty < \mu < \infty$ and $\sigma > 0$.

[ISO 3534-1:2006, definition 2.50]

NOTE 1 The normal distribution is one of the most widely used probability distributions in applied statistics. Owing to the shape of the density function, it is informally referred to as the 'bell-shaped' curve. Aside from serving as a model for random phenomena, it arises as the limiting distribution of averages. As a reference distribution in statistics, it is widely used to assess the unusualness of experimental outcomes.

NOTE 2 The location parameter μ is the mean and the scale parameter σ is the standard deviation of the normal distribution.

B.1.1.9
parametric method
any statistical method which makes an assumption about the form of probability distribution describing the variability of the sampled population

NOTE For example, the method might assume that the distribution is lognormal.

B.1.1.10
skewness
measure of the asymmetry of a population – i.e. the degree to which values extend further on one side of the median than the other

B.1.1.11
standard normal deviate
values corresponding to specified cumulative proportions of a standard normal distribution

NOTE A normal distribution has a mean of 0 and standard deviation of 1.

Examples are:

Cumulative probability	Standard normal deviate
0,05	-1,645
0,50	0,000
0,95	+1,645
0,975	+1,960

B.1.1.12
stratification
division of a population into mutually exclusive and exhaustive sub-populations (called strata), which are thought to be more homogeneous with respect to the characteristics investigated than the total population

B.1.1.13
variance
measure of dispersion, which is the sum of the squared deviations of observations from their average divided by one less than the number of observations

B.1.2 Symbols

The following symbols are used in Annex B:

n = total number of samples or observations

x_i = i-th sample value (with i running from 1 to n)

$x(i)$ = i-th ranked value – that is, the i-th value after sorting the n values into increasing order

μ = population mean

$\bar{x}$ = sample mean

σ = population standard deviation

s = estimated standard deviation

u_p = standard normal deviate corresponding to cumulative probability p

χ_p^2 = chi-squared deviate corresponding to cumulative probability p

X_P = population P-percentile

X_P = estimated P-percentile

SE(z) = standard error of the statistic z

B(r;n,p) = binomial probability that exactly r out of n random samples have a particular characteristic of interest, when the proportion of the entire population having this characteristic is p

CumB(r;n,p) = cumulative binomial probability that up to r out of n random samples have a particular characteristic of interest, when the proportion of the entire population having this characteristic is p

B.2 Probability distributions

B.2.1 General

The 'probability distribution' is a statistical term used to describe the relative frequencies with which different values arise in a given population. The reliability of a testing programme can be improved if the form of the underlying distribution is known (or can reasonably be assumed).

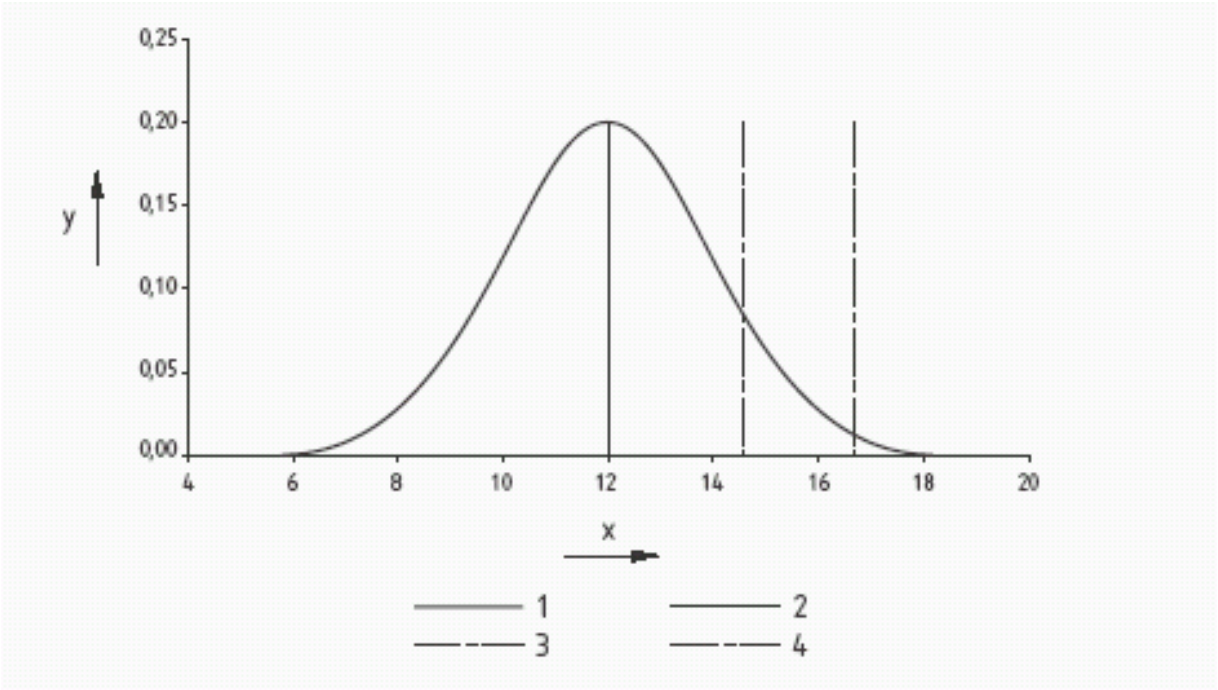
Three distributions of particular relevance to testing programmes are described in the following clauses. It is to be noted that in practice the actual distribution of measurements can be much more complex and as a result differ very much from the three distributions described here.

NOTE For example, bimodal distributions are encountered often due to the fact that two mechanisms contribute to the measured characteristics.

B.2.2 Normal distribution

The probability distribution used most widely in statistics is the normal distribution. This has a characteristic 'bell' shape, and is defined by two quantities or 'parameters': the mean (which fixes the centre of the distribution), and the standard deviation (which determines the degree of spread). These and other statistical parameters are discussed further in B.3.

Figure B.1 shows an example of a normal distribution with mean 12 and standard deviation 2. A characteristic property of a normally distributed population is that about 68% of its observations fall within a range of ± 1 standard deviation from the mean, and about 95% fall within ± 2 standard deviations. Here, therefore, most of the area under the curve lies in the range $12 \pm 2 \times 2$, namely 8 to 16.



Key

1 – Normal curve	2 – Mean, media
3 – 90 %ile	4 – 99 %ile
X - Concentration	Y – Probability density

Figure B.1 – Example of a normal distribution

The normal distribution is important for two main reasons. One is that many standard statistical test procedures (e.g. t-tests, F-tests) rest on the assumption that the sample values have been drawn from a normal population.

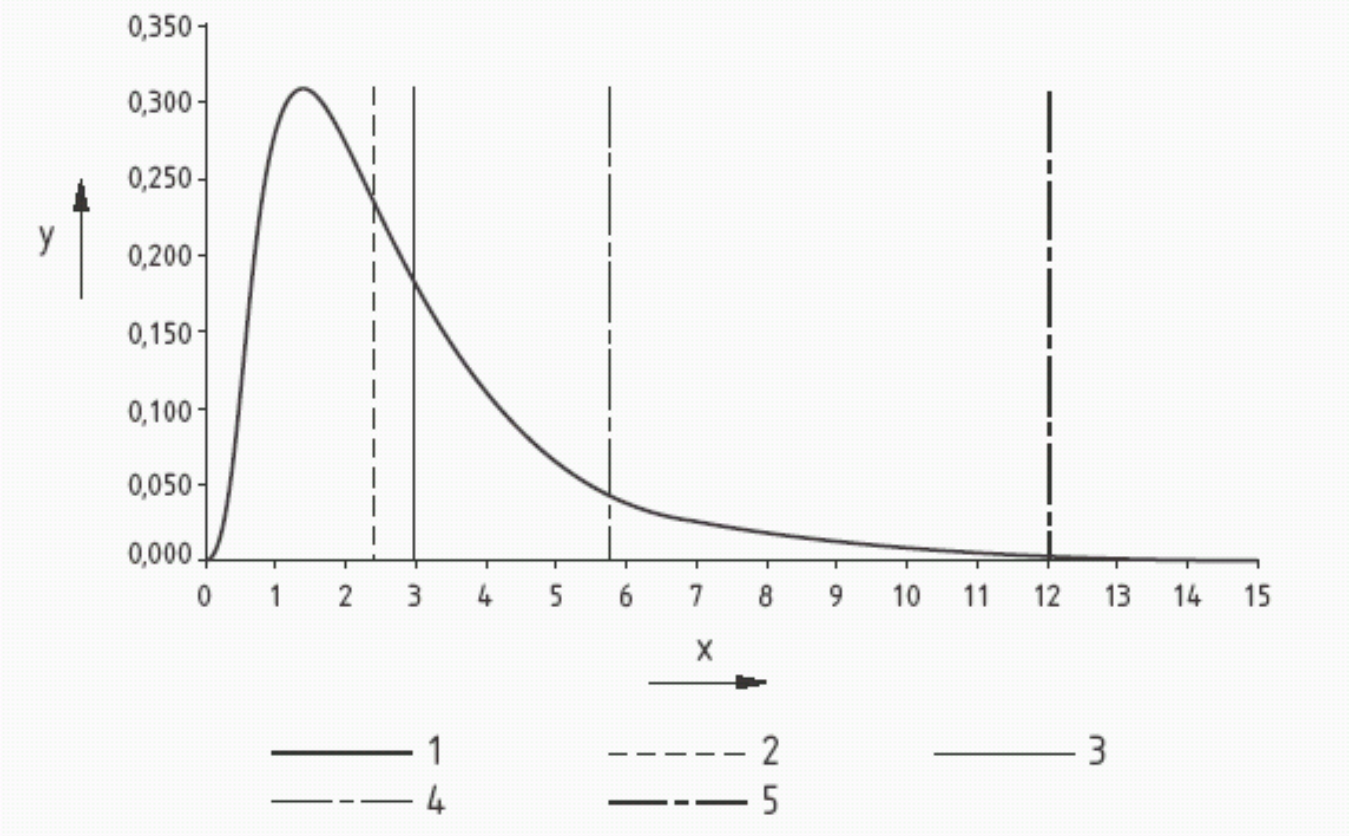
Normality is in general not applicable for the statistical distribution of observations on the composition of a heterogeneous material. It is more common to find a positively skewed distribution, whereby the majority of values are grouped relatively close to zero, but a minority of values form a tail of increasingly larger concentrations. It is easy to see how this can arise: concentrations (or other relevant characteristics of the material) can never be less than zero, but occasional high concentrations can occur.

Such populations are often better described by the lognormal distribution (see B.2.3).

B.2.3 Lognormal distribution

Figure B.2 shows an example of a lognormal distribution with mean 3,0 and standard deviation 2,4 (giving a *relative* standard deviation, or coefficient of variation, of 2,4 / 3,0 = 0,8). The right-hand skewness can clearly be seen: more than 90 percent of the population falls below 6 mg/kg, whilst the greatest 1 percent of the population lies beyond 12 mg/kg.

Relative st.dev = 0,8



Key

1 – Lognormal curve	2 – Median
3 – Mean	4 – 90 %ile
5 – 99 %ile	
X – Concentration	Y – Probability density

Figure B.2 – Example of a lognormal distribution

The lognormal distribution is only a convenient approximation, and cannot always reflect the extreme skewness seen in some types of data. Nevertheless it often provides an acceptable assumption, especially where the purpose of the sampling is to estimate mean concentrations.

There is also a practical advantage in assuming a population to be lognormally distributed, in that after logarithmic transformation the values become normally distributed. Although this makes statistical analysis more straightforward, and in particular allows methods based on standard normal theory to be used, transposition of statistical characteristics calculated on the logarithmically transposed data is not in all cases allowed without loosing the statistical correctness of the estimate.

B.2.4 Tests for normality and lognormality

Statistical techniques are available – known collectively as ‘goodness-of-fit’ methods – for testing whether a given data set could reasonably have come from a specified type of probability distribution, such as the normal or the lognormal. Two fairly well known tests are the chi-squared goodness-of-fit test and the Kolmogirov-Smirnov one-sample test. Although details of how these methods are applied are beyond the scope of this document, they can readily be found in most statistics text books and statistical software packages. Where there is sufficient data (say 50 or more sample values), it is certainly a good idea to examine the reliability of any distributional assumption that is made. However, even without formal statistical testing, a good indication of whether or not the normality assumption is reasonable can be gained simply by

examining a histogram of the data or – even better – looking at a 'normal probability plot'. A quick graphical check of the lognormal assumption can similarly be made by looking at a histogram of the logarithmically transposed values of the data.

B.2.5 Binomial distribution

Some cases will arise where the measurement of interest is not a continuous variable, but is instead an attribute or characteristic of the population that can be either 'present' or 'absent'. In such cases a widely applicable distribution is the 'binomial distribution'. This is defined by two parameters: the number of samples to be taken (n), and the proportion (p) of the population that has the attribute in question. The probability of observing a specific number of samples, r, exhibiting the attribute of interest is given by:

$$B(r;n,p) = \frac{n!}{(n-r)!r!} p^r (1-p)^{n-r} \qquad (B.1)$$

For small values of n, individual binomial probabilities can be evaluated by the straightforward application of this formula. However, it soon becomes a problem for larger values of n. Where binomial or cumulative binomial probabilities are needed, therefore, it is advisable that these are calculated using the statistical functions available in most popular spreadsheet packages.

NOTE Theoretical example:
If a fair coin is tossed 10 times, 'tails' will on average appear 5 times, but because of sampling error the actual number might well be less than or greater than 5. The binomial distribution B(r; 10, 0,5) determines the precise probabilities with which 0, 1, 2, …, 9, or 10 tails will be seen. For example, the probability of getting exactly 5 tails is 24,6%.

Practical application:
Suppose it has been agreed that not more than 50% of tankers containing sludge from a particular operator might contain a particular sort of (easily recognized) material – based on colour or smell, perhaps. From daily on-site verification over 30 days, 19 tankers are identified as containing the material. The binomial distribution can quantify just how unusual it would be to get a proportion as high as 19 / 30 *through sampling error alone*, assuming that the process had truly been complying with the allowed rate of 50% (in this example the probability of getting a result at least as extreme as this is 8%, which is small but not unbelievably so).

B.3 Statistical parameters

B.3.1 General

A key step in planning a testing programme is to specify the statistical parameter that is to be estimated. This is important because the choice generally has a critical bearing on both the type of sampling and the number of samples needed.

NOTE For example, composite sampling is an effective method for estimating mean concentration, but is less appropriate for a percentile- or maximum-related objective.

Except for the equation for the estimation of the statistical parameter itself, a second equation is needed for calculating the statistical uncertainty associated with the estimate. The second of these is a critical piece of information, because it provides the quantitative link between the number of samples and the achievable reliability (i.e. precision and confidence). This is addressed in detail in Annex C.

The following clauses provide equations for each of a number of commonly used parameters.

B.3.2 Notation

B.3.2.1 Mean

The arithmetic mean – usually abbreviated to 'mean' – is the most commonly encountered parameter. It is a very useful measure of the 'central tendency' of a population. An unbiased estimate of the population mean is provided by the sample mean, given by:

$$\bar{x} = \frac{\sum x_i}{n} \quad \text{(B.2)}$$

The uncertainty in $\bar{X}$ is given by:

$$SE(\bar{x}) = \frac{s}{\sqrt{n}} \quad \text{(B.3)}$$

B.3.2.2 Standard deviation

The standard deviation is a widely used measure of the variability of the population. It can be thought of as the root-mean-square of all the units in the population. A (nearly) unbiased estimate of the population standard deviation is calculated as:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{(n-1)}} \quad \text{(B.4)}$$

For normal populations, the uncertainty in s can be assessed using the chi-squared distribution. A C% confidence interval for σ given s can be calculated as:

$$s\sqrt{\frac{(n-1)}{X_{1-p}^{\ 2}}} \text{ to } s\sqrt{\frac{(n-1)}{X_p^{\ 2}}} \quad \text{(B.5)}$$

where p = (1 - C/100)/2.

The square of the standard deviation, s^2, is known as the 'variance'. The variance is of great importance in statistical theory, but is not a practically useful measure for reporting variability as it is not defined in the same dimensions as the observed data.

EXAMPLE Suppose a set of concentrations had a mean of 1,1 mg/kg and a standard deviation of 0,3 mg/kg. The variance would be 0,09 mg^2/kg^2.

B.3.2.3 Coefficient of variation

The variability of a population can also be defined in a non-dimensional manner by the coefficient of variation, CV. An approximately unbiased estimate of the coefficient of variation is given by:

$$CV = \frac{s}{\bar{x}} \quad \text{(B.6)}$$

The uncertainty in CV can be quantified for normal populations, but this information is not required for the present applications.

The coefficient of variation is particularly useful when the variability of different populations is to be compared. For many types of material, it is found that the standard deviation of a constituent tends to increase in proportion with its mean. Thus the *relative* standard deviation – i.e. the CV – is approximately constant, and so this forms a good basis for comparison.

B.3.2.4 Percentiles

B.3.2.4.1 General

The P-percentile of a population is that value below which P % of the population lays.

EXAMPLE In Figure B.1, the 90-percentile has a value of about 14,6 mg/kg. This means that 90% of the population is less than or equal to 14,6 mg/kg; equivalently, 10% of the population lies above 14,6 mg/kg.

Depending on what information is available about the underlying probability distribution, percentiles can be estimated in a variety of different ways, which will result in different estimates for the same percentile. Three methods to estimate a percentile are described in B.3.2.4.2, B.3.2.4.3 and B.3.2.4.4. Given the variety of methods to estimate the percentiles and the differences between these estimates, it is important to specify how percentiles are calculated.

B.3.2.4.2 Percentiles assuming normality

The P-percentile is defined as $\mu + u_p \sigma$,

where p = P/100.

NOTE Standard normal deviates u_p for various values of p are as follows:

P	1	5	10	50	75	90	95	97,5
p	0,01	0,050	0,1	0,5	0,75	0,9	0,95	0,975
u_p	-2,326	-1,645	-1,282	0,000	0,675	1,282	1,645	1,960

For example, the 95-percentile is $\mu + 1{,}645\sigma$, and the 1-percentile is $\mu - 2{,}326\sigma$.

An (almost) unbiased estimate of the P-percentile is given by:

$$X_P = \overline{x} + u_p s \qquad \text{(B.7)}$$

where p = P/100.

An approximate equation for the uncertainty in X_P is:

$$SE(X_P) = s\sqrt{\frac{1}{n} + \frac{u_p^{\,2}}{2(n-1)}} \qquad \text{(B.8)}$$

B.3.2.4.3 Percentiles assuming lognormality

B.3.2.4.2 applies equally to the case of lognormally distributed data, with the following adjustments:

a) the standard deviation s refers to the log-transformed data (it being immaterial whether base-10 or base-e is used);

b) at the end of the calculation X_P, the estimate of the P-percentile, should finally be antilogged to return to the unlogged domain.

B.3.2.4.4 Percentiles – non-parametric approach

If nothing can reliably be assumed about the probability distribution, a ‘non-parametric’ method is suggested. This is somewhat less precise than a parametric method – such as those in the preceding clauses – but is clearly a safer option when the parametric approach cannot be relied upon.

There are numerous slight variants of the non-parametric approach. The one proposed here is the so-called ‘Weibull’ convention, whereby the P-percentile is estimated as follows:

$X_P = X(r)$, where $r = (P/100)(n+1)$ (B.9)

If r is not an exact integer, linear interpolation should be used as follows:

$X_P = (1-d)X(s) + dX(s+1)$ (B.10)

where:

s = integer part of r, and

d = r - s

The concept of standard error is less appropriate for non-parametric methods. Instead, the uncertainty in X_P can be quantified by a conservative confidence interval {$X(r_1)$ to $X(r_2)$}, where r_1 and r_2 are defined by the following cumulative binomial expressions:

r_1 is the largest integer satisfying the condition $CumB(r_1-1; n,p) \leq (1 - C/100)/2$, and

r_2 is the smallest integer satisfying the condition $CumB(r_2-1; n,p) \geq 1 - (1 - C/100)/2$.

NOTE The resulting interval will in general have a confidence coefficient rather larger than C % because of the discrete nature of binomial probabilities.

EXAMPLE Suppose it is required to estimate the 80-percentile cadmium concentration from 39 random samples taken from a dewatered sludge stream, together with a 90% confidence interval.

(1) By the Weibull method, $X_{80} = X(r)$, where $r = (80/100)(39+1) = 32$. Thus X_{80} is estimated by X(32), the ordered sample value with rank 32 (or, equivalently, the 8th largest value).

(2) C = 90%, and so the conditions for r_1 and r_2 are:

$CumB(r_1-1; 39, 0,8) \leq 0,05$, and $CumB(r_2-1; 39, 0,8) \geq 0,95$.

Using appropriate software, we find by experimentation that:

CumB(26; 39, 0,8) = 0.0355 and CumB(35; 39, 0,8) = 0,9668.

Thus the interval X(27) to X(36) – that is, the interval from the 13th biggest to the 4th biggest sample value – provides a conservative 90% confidence interval for the true 80-percentile cadmium concentration (the actual confidence coefficient is 0,9668 – 0,0355 = 0,931, or 93,1%).

B.3.2.5 Maximum

The population maximum should never be used as the desired statistical parameter (except in the unlikely event of the sampling being of very high frequency). This is because no reliable estimate of the maximum can ever be obtained from a set of sample values. The sample maximum will always be an under-estimate of the population maximum, and furthermore there is no straightforward method available for quantifying the extent of that bias.

Where the primary objective is concerned with ‘worst case’ values, the objective should be recast in terms of a suitably high percentile – say the 99-percentile. The methods described in B.3.2.4 can then be applied.

B.3.2.6 Percentage compliance with a given limit

B.3.2.6.1 General

The primary sampling objective often relates to the percentage of a population that complies with a specific limit (e.g. a target or intervention value). This is especially true for compliance testing and on-site verification.

As with percentile-type objectives, both parametric and non-parametric approaches can be taken. To contrast the two approaches, imagine that the limit L should be complied with for P% of the time or better.

B.3.2.6.2 Percentage compliance – parametric approach

Using the parametric approach, the P-percentile would be estimated assuming a particular distribution (e.g. normal), and the resulting estimate X_P would be compared with L. The statistical uncertainty in the compliance result would then be assessed using the quantity $SE(X_P)$.

The parametric approach should not be used unless there is reliable information about the nature of the underlying distribution, because of the confusion that can be caused whenever the parametric estimate differs markedly from the non-parametric compliance figure – i.e. the simple pass rate calculated directly from the data. Moreover, the details of the statistical method are beyond the scope of this document (even in the case where normality can be assumed), and specialist statistical advice should be sought for its application.

B.3.2.6.3 Percentage compliance – non-parametric approach

By the non-parametric approach, the quantity r – the number of sample values $\leq$ L – is first calculated. The sample compliance 100(r/n)% can then be determined. The advantage now is that 100(r/n) is binomially distributed (irrespective of the distribution followed by the original samples), and so the statistical uncertainty in the compliance result can be assessed without the need for any distributional assumptions about the population. Specifically, a C% confidence interval for the true population compliance is given by [$100p_{LO}$ to $100p_{UP}$], where:

p_{LO} is chosen so that $1 - CumB(r-1; p_{LO}, n) = (1 - C/100)/2$, and

p_{UP} is chosen so that $CumB(r; p_{UP}, n) = (100 - C)/2$.

NOTE Although the definition of the limit with which the observations are to be compared falls outside the scope of this standard, it is important to realize that the (often implicit) statement that 'no observation may exceed the limit' is statistically unusable. It implies that not even one single unit of the population (at the investigated scale, see also Annex A) might have a concentration above that limit. In order to test this hypothesis, it would be necessary to test the entire population at the predefined scale!

However, an almost equivalent but statistically 'coherent' level of protection can be obtained by requiring that 99 % (or even 99,9 %) of the population at the defined scale, rather than 100%, should comply with the limit.

EXAMPLE Suppose that bags of pelletized sewage sludge leaving a processing plant are required to have pH > 6. It is decided routinely to submit half of the outgoing bags to on-site verification. After several months, 300 out of the 600 bags produced have been checked, and all are satisfactory. Even from evidence as strong as this, it is impossible to say with any confidence that all 600 bags were satisfactory. However, what can be said with 95% confidence is that at least 99 % of the bags will comply because Cum P(0; 0,99, 300) = 0,049.

Annex C
(informative)
Calculating the required number of increments and samples

C.1 Symbols

The following symbols are used in Annex C:

n = total number of samples or observations

m = number of increments per composite sample

μ = population mean

u_p = standard normal deviate corresponding to cumulative probability p

χ_p^2 = chi-squared deviate corresponding to cumulative probability p

X_P = population P-percentile

SE(z) = standard error of the statistic z

σ_w = standard deviation of local (i.e. within-composite) spatial variation

σ_b = standard deviation of between-composites spatial and/or temporal variation

σ_s = standard deviation of total spatial and/or temporal variation (= $\sqrt{[\sigma_w^2 + \sigma_b^2]}$)

NOTE In cases where composite sampling is not being considered, spot samples can be thought of as composite samples with just a single increment, and so the 'within-composite' standard deviation becomes zero, and the 'between-composites' standard deviation becomes the 'between-spots' standard deviation.

σ_e = standard deviation of analytical error

C = desired confidence level (%)

a = cumulative probability related to the desired confidence level

d = desired precision

C.2 Estimating a mean concentration

C.2.1 Using composite samples

The standard error of the mean is given by:

$$SE(mean) = \sqrt{[(\sigma_w^2/m + \sigma_b^2 + \sigma_e^2)/n]} \quad (C.1)$$

Thus for a given value of m, and assuming normality, the number of composites required to achieve the desired precision (d) and confidence (C), as specified by the user, is given approximately by:

$$n = (u_a/d)^2(\sigma_w^2/m + \sigma_b^2 + \sigma_e^2) \quad (C.2)$$

where a = 1 - (1 - C/100)/2. Alternatively, equation C.2 can be re-written to determine the number of increments (m) needed per composite sample if n, the total number of composite samples, has been set in advance. Thus:

$$m = \sigma_w^2/[\, n(d/u_a)^2 - \sigma_b^2 - \sigma_e^2\,] \qquad (C.3)$$

NOTE It might be desirable – especially for on-site verification – to plan to take only a single composite sample. Provided $\sigma_b^2 + \sigma_e^2$ is sufficiently small, this can be achieved by setting n equal to 1 in equation C.3.

In practice, the true standard deviations are unknown and so estimates have to be used. In some cases it might be appropriate to use the values obtained from the past analysis of sample data from similar investigations. Otherwise the estimates should, where possible, be obtained from a preliminary pilot study.

EXAMPLE 1 Suppose that:

— estimates of σ_w, σ_b and σ_e are 4, 2 and 0,5 mg/kg;

— 10 increments are to be taken per composite (i.e. m = 10); and

— the mean is required to be estimated to a precision of d = 1 mg/kg with 90% confidence.

For C = 90, a = 1 - (1 - 90/100)/2 = 0,95, and so u_a = 1,65.

From equation C.2, $n = (1,65)^2(16/10 + 4 + 0,25) = 15,9$.

Thus about 16 composite samples would be needed to produce a mean to the required reliability.

To decide on the most appropriate value of m it is necessary to consider the relative costs of sampling and analysis. Suppose that the sampling cost per increment is A, and the analysis cost per sample is B. The total cost TC is accordingly given by:

$$TC = (Am + B)n \qquad (C.4)$$

Thus, using equation C.2 with various trial values of m it is possible to find the combination of m and n which minimizes TC.

EXAMPLE 2 Continuing with the earlier example, suppose that:

— values of m ranging from 1 to 20 are considered; and

— B/A = 30 – that is, a sample analysis is 30 times more expensive than the cost of sampling an increment.

Figure C.1 shows the n value given by equation C.2 for each trial value of m. Figure C.2 shows the corresponding values of the total sampling cost TC (in arbitrary units). It is apparent that the optimum number of increments per composite sample is about 6.

C.2.2 Using individual samples

The standard error of the mean is given by:

$$SE(mean) = \sqrt{[(\sigma_s^2 + \sigma_e^2)/n]} \qquad (C.5)$$

Thus the number of samples required to achieve the desired precision (d) and confidence (C), as specified by the user, is given approximately by:

$$n = (u_a/d)^2(\sigma_s^2 + \sigma_e^2) \qquad (C.6)$$

where a = 1 - (1 - C/100)/2.

NOTE Individual sampling can be thought of as composite sampling with just one increment per composite. Thus the results of the previous section apply to the case of spot sampling by substituting m = 1 and replacing $\sigma_w^2 + \sigma_b^2$ with σ_s^2.

In practice, the true standard deviations are unknown and so estimates have to be used. In some cases it might be appropriate to use the values obtained from the past analysis of sample data from similar investigations. Otherwise the estimates should where possible be obtained from a preliminary pilot study.

EXAMPLE Suppose that:

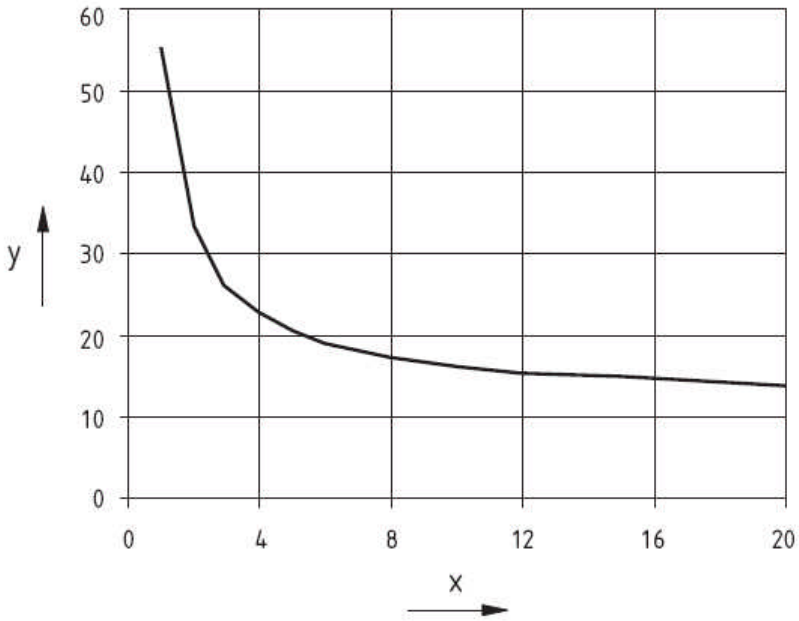
— estimates of σ_s and σ_e are 4,5 and 0,5 mg/kg; and

— the mean is required to be estimated to a precision of d = 2 mg/kg with 90% confidence.

For C = 90, a = 1 - (1 - 90/100)/2 = 0,95, and so u_a = 1,65.

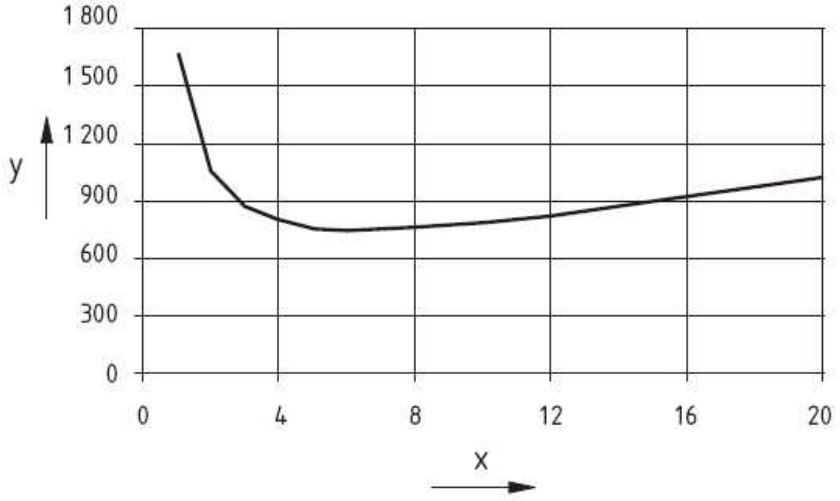
From equation C.6, n = $(0,825)^2(20,25 + 0,25)$ = 13,9.

Thus about 14 individual samples would be needed to produce a mean to the required reliability.



Key
X Number of increments, m
Y Required no. of composites, n

Figure C.1 – Illustration of the relationships between m, n and TC (see text for details) – Samples needed to achieve specified precision and confidence



Key
X Number of increments, m
Y Total cost of sampling

Figure C.2 – Illustration of the relationships between m, n and TC (see text for details) – Cost of sampling in relation to number of increments per composite sample

C.3 Estimating a standard deviation

The following approach is applicable when the population can be assumed to be normally distributed. Even for non-normal populations, however, the method is useful as a rough approximation.

Confidence intervals for σ can be calculated using the equation given in B.3.2.2. For a given choice of confidence C, this can be evaluated for a range of trial n values, and this will identify the number of samples that provides the required precision.

EXAMPLE Suppose it is required to estimate the standard deviation to a precision of 20 % with 90 % confidence. For 90 % confidence, the lower and upper p values are = $(1 \pm C/100)/2$ = 0,05 and 0,95. With the help of statistical tables of the χ^2 distribution at the p = 0,05 and 0,95 points, the following table can be constructed:

Table C.1 – 90 % confidence limits for σ/s for various numbers of samples

Number of samples	Lower 90 % confidence limit for σ/s	Upper 90 % confidence limit for σ/s
n	$\sqrt{[(n-1)/\chi^2]}$ (p = 0,05)	$\sqrt{[(n-1)/\chi^2]}$ (p = 0,95)
20	0,79	1,37
30	0,83	1,28
40	0,85	1,23
50	0,86	1,20
60	0,87	1,18
70	0,88	1,16
80	0,89	1,15
90	0,89	1,14
100	0,90	1,13
120	0,90	1,12

150	0,91	1,11
200	0,92	1,09

By inspection it can be seen that with 50 samples, the lower and upper confidence limits are 0,86 and 1,20. That is, the population standard deviation σ might be 14% below or 20% above s, the observed standard deviation. Note that the interval is not symmetrical. Thus, at the 90% confidence level, a precision of 20% or better will be achieved by a standard deviation calculated from 50 random samples.

C.4 Estimating a percentile

C.4.1 Assuming normality

The standard error of the P-percentile X_P is given by:

$$SE(X_P) = \sigma\sqrt{\frac{1}{n} + \frac{u_p^{\,2}}{2(n-1)}} \qquad (C.7)$$

where p = P/100, and $\sigma = \sqrt{(\sigma_s^{\,2} + \sigma_e^{\,2})}$.

Thus the number of samples required to achieve the desired precision and confidence is given approximately by:

$$n = (u_a s/d)^2(1 + u_p^{\,2}/2) \qquad (C.8)$$

where a = 1 - (1 - C/100)/2, and s is an estimate of σ.

EXAMPLE Suppose that

— σ is estimated by s = 3,5 mg/kg; and

— the 95-percentile is required to be estimated to a precision of d = 1,46 mg/kg with 90% confidence.

For the 95-percentile, p = 0,95 and so u_p = 1,65.

For C = 90, a = 1 - (1 - 90/100)/2 = 0,95, and so u_a = 1,65.

Thus from equation C.8, $n = (1{,}65\times3{,}5/1{,}46)^2(1 + 1{,}65^2/2) = 36{,}9$.

Thus about 37 samples would be needed for the 90-percentile to be estimated to the required reliability.

C.4.2 Non-parametric approach

For determining the precision achievable by a non-parametric approach, there is no direct expression available corresponding to the one given above for the normal case. As a rough approximation, however, the equation given in C.4.1 can still be used, but with an additional multiplicative factor of 1,3 applied to represent the poorer precision typically attained by the non-parametric rather than the normal-based approach.

Alternatively, exact results can be obtained using the following more time-consuming approach. The first step is to select a trial number of samples and desired confidence level, C. The methodology described in B.3.2.4.4 for calculating C % confidence intervals around non-parametric percentile estimates is then applied. This should be repeated for different trial sample numbers. The various confidence intervals will be expressed as ranked values, but these can be converted into equivalent actual measurements as long as a suitable historical data set is available. These trial calculations will give an indication of the precision that can typically be achieved at C% confidence for various numbers of samples; and from this an appropriate choice can be made.

EXAMPLE Suppose that the 80-percentile cadmium concentration from a particular sludge stream is required to be estimated to a precision of d = 15 mg/kg with 90% confidence.

Select n = 39 as the trial number of samples.

From B.3.2.4.4, a conservative 90 % confidence interval is provided by the interval X(27) to X(36).

Past cadmium data is available for this sludge stream. From a set of 39 values taken at random from this data, the 15 highest values are: 12, 12, 13, 15, 17, 20, 20, 25, 26, 31, 31, 35, 36, 40, 55 mg/kg.

Thus the 27th and 36th ranked values are 13 and 35 mg/kg and so the expected precision is (35 - 13)/2 = 11 mg/kg. This is better than required, and so a lower trial value of n is selected.

Select n = 29 as the new trial number of samples.

From B.3.2.4.4, a conservative 90% confidence interval is provided by the interval X(20) to X(28).

From a set of 29 values taken at random from the historical data, the 12 highest values are: 10, 12, 12, 15, 20, 20, 25, 26, 31, 35, 40, 55 mg/kg.

Thus the 20th and 28th ranked values are 12 and 40 mg/kg and so the expected precision is (40 - 12)/2 = 14 mg/kg. This is adequately close to the required precision.

About 29 samples would therefore be needed for the 80-percentile to be estimated to the required reliability.

C.5 Estimating a percentage compliance with a given limit

The approach here is similar to that described in C.4.2. First the desired confidence level, C, is chosen. Then, for each of a range of trial sample numbers, the C% confidence interval for the true percent compliance is calculated using the methodology described in B.3.2.6. The resulting set of confidence intervals shows the quantitative link between achievable precision and samples taken, and hence provides a rational basis for arriving at an acceptable compromise.

EXAMPLE Suppose that:

— the percentage of material meeting a particular cadmium concentration limit is thought to be about 80%;

— this percentage has to be estimated to a precision of 10% with 90% confidence; and

— nothing is known about the statistical nature of the cadmium distribution

Select a trial number of samples of n = 20, and suppose that 16 samples meet the required cadmium limit (that is, the observed compliance rate is 80%).

Using the non-parametric binomial method described in B.3.2.6.3, calculate a 90 % confidence interval for the true compliance percentage. This is 71,7% - 98,2%, giving a precision of about 13 %. Thus a greater number of samples is needed.

Select a trial number of samples of n = 40, and suppose that 32 samples meet the required cadmium limit (to keep the observed compliance rate at 80%).

Using B.3.2.6.3, calculate a 90 % confidence interval for the true compliance percentage. This is 78,6 % - 96,5 %, giving a precision of about 9 %. This is adequately close to the required precision.

About 40 samples would therefore be needed for the compliance percentage to be estimated to the required reliability.

Annex D
(informative)
Minimum increment and sample size mass (mass / volume)

D.1 General

The sampling plan should contain specific instructions on the type of samples to be taken, the size of increments and/or samples, the number of increments and/or samples to be taken and, when relevant, the number of increments that should be put together in a composite sample.

D.2 Estimation of increment and sample size

D.2.1 General

As mentioned in 5.4, a key feature of probabilistic sampling is that all parts of the population have the chance of being part of the sample. For the sampling of granular material, this has an effect on the scale (volume or mass) of both increments and samples. This paragraph and subsequent sub-paragraphs show how the increment and sample size should be determined according to the following steps:

1) determination of the minimum increment size;

2) determination of the minimum sample size;

3) determination of the number of increments and/or samples;

4) calculation of the actual increment and/or sample size.

D.2.2 Determination of the minimum increment size

The minimum increment size when sampling from a static batch should meet the following requirements:

— the actual width, height and length of the sampling equipment should be at least equal to three times the 'maximum' particle size (D_{95}) of the material to be sampled in the case of materials with a maximum particle size (D_{95}) of at least 3 mm;

— the actual width, height and length of the sampling equipment should be at least equal to 10 mm in the case of materials with a maximum particle size (D_{95}) of less than 3 mm.

EXAMPLE 1 $D_{95} \geq 3$mm

If the maximum particle size is at least 3 mm and the width, height and length of the increment are chosen to be equal to three times the maximum particle size (D_{95}), then the following formula applies to the mass of the minimum increment size:

$$M_{inc} = 10^{-9}\rho(3D_{95})^3 = 2{,}7\times10^{-8}\rho D_{95}^3 \quad (D.1)$$

where:

M_{inc} = mass of minimum increment size, in kg,

D_{95} = the 95-percentile particle size, in mm, and

ρ = the bulk density of the material, in kg/m^3.

Moreover, the mass of the maximum particle is $(4/3)\pi\rho10^{-9}(D_{95}/2)^3 = 5{,}2\times10^{-10}\rho D_{95}^3$.

Thus the quantity (mass of increment)/(mass of maximum particle) = 270/5,2 = 51,6. In other words, the mass of the minimum increment should be about 50 times that of a maximum particle (95-percentile of the particle size distribution).

EXAMPLE 2 D_{95} <3mm

In the case of materials with a maximum particle size (D_{95}) of less than 3 mm, the following formula applies to the mass of the minimum increment size:

$$M_{inc} = 1 \times 10^{-6} \rho \quad \text{(D.2)}$$

where:

M_{inc} = the mass of the minimum increment size, in kg, and

ρ = the bulk density of the material, in kg/m^3.

D.2.3 Determination of the minimum sample size

The minimum sample size to be applied to the material in question is given by:

$$M_{sam} = \frac{1}{6} \pi \times (D_{95})^3 \times \rho \times g \times \frac{(1-p)}{CV^2 \times p} \quad \text{(D.3)}$$

where:

M_{sam} = the mass of the sample in g;

D_{95} = the 'maximum' particle size (defined as the 95-percentile), in cm;

ρ = the specific mass of the particles in the material, in g/cm^3;

g = the correction factor for the particle size distribution of the material to be sampled;

p = is the fraction of the particles with a specific characteristic (m/m);

CV = the desired coefficient of variation caused by the fundamental error.

Note that this calculation results only in a rough estimate of the minimum sample size. The estimate however is precise enough to know the order of magnitude of the sample size. Two, partly related, aspects determine the correctness of the estimate: the quality of the assumptions made for the parameters in the formula (thus how correct are the estimates) and the correctness of the formula itself for non-spherical particles. As the aim is to obtain a (rough) estimate of the minimum sample size, the formula can also be used for non-spherical (e.g. irregularly shaped materials) or even non-granular materials.

NOTE

1) The variables in the formula for the estimation of the minimum sample size are expressed in CGS units for practical reasons.

2) The minimum sample size is directly related to the desired coefficient of variation of the fundamental error (CV) and to the size of fraction of the particles with the characteristic to be determined (p). The result is derived from binomial sampling theory as follows. Suppose n samples are taken from the material. The standard error of the observed proportion of particles with the characteristic of interest is $\sqrt{[p(1-p)/n]}$, and so the coefficient of variation CV is given by:

$CV^2 = (1-p)/(pn)$.

Thus to achieve an adequately small value of CV, the value of n should be:

$n = (1-p)/[CV^2 p]$ - which is the final term in the equation for M_{sam}.

3) As the influence of the fundamental variability (see A.2) should be low, a well accepted value for the coefficient of variation due to the fundamental variability is 0,1.

4) The actual value of p, the fraction of particles with a certain characteristic, depends on the material to be sampled and the substances in it to be determined. Knowledge of the material consistency is required in order to determine this value.

5) The formula for estimating the minimum sample size is derived for spherical particles of diameter d, and so is only an approximation for non-spherical particles.

6) The following applies for the correction factor for the particle size distribution (g):

Broad particle size distribution:	$D_{95}/D_{05} > 4$	g = 0,25
Medium particle size distribution:	$2 < D_{95}/D_{05} \leq 4$	g = 0,50
Narrow particle size distribution:	$1 < D_{95}/D_{05} \leq 2$	g = 0,75
Uniform particles:	$D_{95}/D_{05} = 1$	g = 1,00

where D_{05} = the 'minimum' particle size (defined as the 5-percentile of the particle size distribution).

7) For the sampling of fine granular material with a broad particle size distribution (e.g. soil), the following default values can be used for the factors in the formula:

ρ = 2,6 g/cm^3

g = 0,25

p = 0,02

D.3 Determination of the number of increments and/or samples

The number of increments and/or samples is directly related to the objective of the testing programme (Clause 4), the variability of the material to be sampled (5.3), and the desired precision and confidence (5.7). Reliable information on variability is commonly unavailable – in which case it is not possible to fulfil the exact requirements for precision and confidence without carrying out a preliminary sampling investigation.

From these specified inputs, the number of increments (where relevant) and samples can be calculated using the equations given in Annex C.

D.4 Calculation of the actual increment and/or sample size

D.4.1 General

On the basis of the relationship between the minimum increment size (D.2.2), the minimum sample size (D.2.3) and the number of increments to be included per composite sample (D.3), the actual increment size and the actual sample size should be determined according to the following rules.

D.4.2 Taking individual samples

Where composite sampling is not being considered, the question of increment size will in most cases be irrelevant as the mass of the minimum sample size (D.2.3) will exceed the mass of the minimum increment size. When the amount of material necessary for the analysis exceeds the mass of the minimum sample size, the actual sample size should of course be sufficient for the analysis.

D.4.3 Composite sampling

Where composite sampling is to be undertaken, there is a possible conflict between:

a) the previously calculated minimum values for increment size (D.2.2) and sample size (D.2.3), and

b) the planned number of increments, m (Annex C).

Such conflict should be resolved as follows:

— if m increments of minimum size amount to *less than* the required minimum sample size, then the increment size needs to be increased accordingly so that m *actual* increments will produce an adequately large composite sample;

— conversely, if m increments of minimum size amount to *more than* the required minimum sample size, then this larger quantity defines the actual sample size.

Annex E
(informative)
Example sampling scenarios

E.1 Introduction

A number of examples of sampling scenarios have been developed to illustrate some approaches that might be required for sludge, treated biowaste, and soils in the landscape.

E.2 Example 1: Sludge producer to carry out a basic characterisation of an agricultural field to determine if sludge can be applied

An assessment is required of an agricultural field to determine the basic soil characteristics to determine the levels of sewage sludge to be applied. Composite samples are to be collected across the field for basic characterization. In practice, a full range of parameters would need to be assessed, but to simplify this example the calculations will be undertaken for just one parameter – phosphate. This represents only one part of the complete characterization programme that could be required by the regulator. Other tests might include an assessment of pH, metal concentrations and other nutrients such as nitrate.

Specify the objective of the testing programme		
1	**Specify the objective in terms of the overall population**	Sludge producer to determine the phosphate requirement in an agricultural field.
Develop the technical goals from the objective		
2	**Define the population to be sampled**	Population: The entire field to which sludge might be applied. Sub-population: The entire field or part of the field that is not more than 5 hectares in area and is managed for the same purpose. NOTE The boundaries of the population and sub-population should be determined using information from maps and in consultation with the farmer.
3	**Assess variability**	Spatial variability is assumed to be approximately equal to variability reported for similar soil types and agricultural management schemes. These soils are characterized by a mean concentration of 40 mg P/kg and a standard deviation of 22 mg/kg. Analytical error known is known from AQC records to be 3 mg/kg.
4	**Select the sampling approach**	Probabilistic sampling is feasible because the whole area can be accessed.
5	**Identify the scale**	The mean concentration for the field should be determined; the scale is the sub-population.
6	**Identify the required statistical approach**	Mean concentration of phosphate expressed as mg P/kg.
7	**Choose the desired reliability**	Required parameter (i.e. 90-percentile) to be estimated to a precision of 20 % with 90 % confidence.
Determine the practical instructions		
8	**Choose the sampling pattern**	Sample across the whole area.
9	**Determine increment /sample size**	Increment samples to be collected through the full soil profile to a depth of 25 cm.
10	**Determine the use of composite or individual samples**	Composite sampling.
11	**Determine required number of samples**	From Annex C, formula for number of increments in a composite is: $m = \sigma_w^2/[\, n(d/u_a)^2 - \sigma_b^2 - \sigma_e^2\,]$, where $a = 1-(1-C/100)/2$. Here d = 20% of 40 = 8 mg/kg; n = 1; and estimates of σ_w and σ_e are 22 and 2 mg/kg. Also, σ_b can be set to zero as the composite sample is being drawn from the entire sub-population. Finally 90% confidence is required, thus a = 0,95, and u_a = 1,65. This gives $m = 484/[(8/1,65)^2 – 4] = 24,8$. Thus a single composite sample

		comprising 25 increments will provide the required precision.
12	**Define statistical elements of the sampling plan**	From a plan of the field, select 25 locations spaced roughly equally over the whole area. Collect one increment from each area and mix the increments into a single composite sample.

E.3 Example 2: Sludge producer to carry out compliance testing of the copper concentration of thermally treated sewage sludge granules in large bags

The sludge producer has previously carried out characterization to determine the extent of variation of bag-to-bag variation in mean Cu concentration. This was characterized by the standard deviation, s. The sludge producer wishes to implement a compliance testing scheme. For this an upper 99% control limit defined by U = L + 2,33s, where L is the permitted long-term mean Cu concentration. About once a week a bag is to be selected at random and the mean Cu concentration determined. If this falls below U, the process is judged to be 'in control'. If a value exceeds U, this alerts the producer to the possibility that the process might have slipped out of control, for example due to an accidental discharge to the sewerage system.

Specify the objective of the testing programme		
1	**Specify the objective in terms of the overall population**	Sludge producer to carry out compliance testing of Cu content of thermally dried sewage sludge granules held in large bags against a permitted mean limit of 1000 mg/kg.
Develop the technical goals from the objective		
2	**Define the population to be sampled**	The sludge granules in large bags over a 3-month period. NOTE All bags produced will be equally accessible, and so available for sampling.
3	**Assess variability**	From previous basic characterization, bag-to-bag variation in mean Cu concentration has an estimated standard deviation of 168 mg/kg.
4	**Select the sampling approach**	Probabilistic sampling is feasible because of good access to bags.
5	**Identify the scale**	Information is required on the between-bag scale; the scale is the volume of a bag.
6	**Identify the required statistical approach**	Mean Cu concentration.
7	**Choose the desired reliability**	Confidence is to be 99% that a non-compliant bag represents a genuine increase in long-term mean Cu. For 99% confidence, corresponding standard normal deviate is 2,33s thus control limit is set at L + 2,33s, namely 100 + 2,33 x 168 = 391,4. In practice it is convenient to round this up to 400 mg/kg as a convenient pass/fail criterion.
Determine the practical instructions		
8	**Choose the sampling pattern**	Select bags at a systematic frequency of 1 in 245 from the production stream. NOTE bags of granules are produced at about 30-40 per day. There is no reason to suspect every 245th bag of having any recurring feature that will bias the assessment.
9	**Determine increment /sample size**	Sludge granules adequately mixed, so increment size is adequate. Check made that sample volumes meet requirements for analysis. NOTE With the sampling implement used, six increments provide a satisfactory composite sample.
10	**Determine the use of composite or individual samples**	Experience indicates that six increments taken at progressive depths through the bag will provide a satisfactory composite sample.
11	**Determine required number of samples**	Not relevant for an on-going compliance scheme based on control chart principles.

12	**Define statistical elements of the sampling plan**	Starting with a randomly chosen bag, select every 245th bag thereafter. For each sampled bag, take a composite sample as specified above and determine the average Cu concentration for that bag. Compare with control limit U.

E.4 Example 3: A treated biowaste producer is to test a batch of composted biowaste for compliance with treatment requirements

An assessment is required of the *E. coli* concentration in a heap of composted biowaste to confirm that treatment requirements have been complied with.

Specify the objective of the testing programme		
1	**Specify the objective in terms of the overall population**	Treated biowaste producer to carry out a compliance characterisation of *E. coli* in a heap of composted biowaste. The compliance level is five samples to be tested and none to exceed 1000 CFU/g.
Develop the technical goals from the objective		
2	**Define the population to be sampled**	Population: The heap of composted biowaste.
3	**Assess variability**	Variability on a batch-by-batch basis is likely. Temporal variability not relevant. Substantial spatial variation. Analytical error known from AQC records. Some past data available, for which mean = 2,35 $\log_{10}$ CFU/g and standard deviation = 0,23 $\log_{10}$ CFU/g. Thus historical 95-percentile is approximately mean + 2,81 st.dev = 3,00 $\log_{10}$ CFU/g
4	**Select the sampling approach**	Probabilistic sampling is not feasible because it is not possible to access to the full volume of the compost heap. Individual samples are needed rather than composites because the required parameter is not mean concentration.
5	**Identify the scale**	Not relevant.
6	**Identify the required statistical approach**	No samples of the compost to exceed the set compliance level.
7	**Choose the desired reliability**	Failure should be declared if any of the samples fail to comply with the limit.
Determine the practical instructions		
8	**Choose the sampling pattern**	Samples taken at the end of the compost maturing time.
9	**Determine increment /sample size**	Sufficient sample to complete the analysis.
10	**Determine the use of composite or individual samples**	Analyse each sample separately.
11	**Determine required number of samples**	The number of samples is specified in regulations.
12	**Define statistical elements of the sampling plan**	Take five samples randomly form within the heap.

Bibliography

BS 2846-1:1991, Guide to statistical interpretation of data – Part 1: Routine analysis of quantitative data

EN 1482:1996, Sampling of solid fertilizers and liming materials

EN 5667-13:1997, Water quality – Sampling – Part 13: Guidance on sampling of sludges from sewage and water treatment works

EN 5667-15:1999, Water quality – Sampling – Part 15: Guidance on preservation and handling of sludge and sediment samples

EN 5667-16:1998, Water quality – Sampling – Part 16: Guidance on biotesting of samples

EN 5667-3:1995, Water quality – Sampling – Part 3: Guidance on the preservation and handling of samples

EN 5667-3:2003, Water quality – Sampling – Part 3: Guidance on the preservation and handling of water samples

EN 12579:1999, Soil improvers and growing media – Sampling

EN 12832:1999, Characterization of sludges – Utilization and disposal of sludges – Vocabulary

EN 13040:1999, Soil improvers and growing media – Sample preparation for chemical and physical tests, determination of dry matter content, moisture content and laboratory compacted bulk density

ISO 2602:1980, Statistical interpretation of data – Estimation of the mean: confidence interval

ISO 2854:1976, Statistical interpretation of data – Techniques of estimation and tests relating to means and variances

ISO 3301:1975, Statistical interpretation of data – Comparison of two means in the case of paired observations

ISO 3494:1976, Statistical interpretation of data – Power of tests relating to means and variances

ISO 3534-1, Statistics – Vocabulary and symbols – Probability and general statistical terms

ISO 3534-2, Statistics – Vocabulary and symbols – Statistical quality control

ISO 3534-3, Statistics – Vocabulary and symbols – Design of experiments

ISO 5479:1997, Statistical interpretation of data – Tests for departure from normality

ISO 11074-2:1998, Soil Quality – Part 1: Terminology and classification – Section 1.2: Terms and definitions relating to sampling

ISO 16269-7:2001, Statistical interpretation of data – Median – Part 7: Estimation and confidence intervals

prEN xxxxx: Sludge, treated biowaste, and soils in the landscape – Sampling – Framework for the preparation and application of a sampling plan

prCEN/TR XXXX-2: Sludge, treated biowaste, and soils in the landscape – Sampling – Part 2: Guidance on sampling techniques

prCEN/TR XXXX-3: Sludge, treated biowaste, and soils in the landscape – Sampling – Part 3: Guidance on sub-sampling in the field

prCEN/TR XXXX-4: Sludge, treated biowaste, and soils in the landscape – Sampling – Part 4: Guidance on procedures for sample packaging, storage, preservation, transport and delivery

prCEN/TR XXXX-5: Sludge, treated biowaste, and soils in the landscape – Sampling – Part 5: Guidance on the process of defining the sampling plan