Eurachem

CITAC

Co-Operation on International Traceability in Analytical Chemistry

EURACHEM / CITAC Guide

Measurement uncertainty arising from sampling

A guide to methods and approaches

Produced jointly with
EUROLAB, Nordtest and the
UK RSC Analytical Methods Committee

First Edition 2007

Measurement uncertainty arising from sampling: A guide to methods and approaches

Produced jointly by
Eurachem, EUROLAB, CITAC, Nordtest and the RSC Analytical Methods Committee

Editors

Michael H Ramsey (University of Sussex, UK)

Stephen L R Ellison (LGC, UK)

Composition of the Working Group

Eurachem members

Michael H Ramsey (Chair)	University of Sussex, UK
Stephen L R Ellison (Secretary)	LGC, UK
Paolo de Zorzi	National Environmental Protection Agency, Italy
Pentti Minkkinen	Lappeenranta University of Technology, Finland
Roger Wood	Food Standards Agency, UK
Máire C Walsh	State Chemist for Ireland
Alex Williams	UK
Christian Backman	GTK, Finland
Maria Belli	National Environmental Protection Agency, Italy

EUROLAB members

Manfred Golze	BAM, Germany
Rüdiger Kaus	HS Niederhein, Denmark
Ulrich Kurfürst	FH Fulda, Denmark
Mikael Krysell	Eurofins Denmark

CITAC member

Ilya Kuselman	National Physical Laboratory, Israel

Nordtest representatives

Christian Grøn	DHI – Water & Environment, Denmark
Bertil Magnusson	Technical Research Institute of Sweden
Astrid Nordbotten	Food Safety Authority, Norway

AMC representatives

Michael Thompson	University of London, UK
Jenny Lyn	Food Standards Agency, UK

Observer

Marc Salit	NIST, USA

Acknowledgements

This document has been produced primarily by a joint Eurachem/EUROLAB/CITAC/Nordtest/ AMC Working Group with the composition shown (right) working in the period 2004–2006. The editors are grateful to all these individuals and organisations and to others who have contributed comments, advice and assistance.

Production of this Guide was in part supported under contract with the UK Department of Trade and Industry as part of the National Measurement System Analytical Measurement (VAM) Programme, and also by the UK Food Standards Agency and Environment Agency.

This publication should be cited as:
M H Ramsey and S L R Ellison (eds.) Eurachem/EUROLAB/ CITAC/Nordtest/AMC Guide: *Measurement uncertainty arising from sampling: a guide to methods and approaches Eurachem* (2007). ISBN 978 0 948926 26 6. Available from the Eurachem secretariat

Foreword

Uncertainty of measurement is the most important single parameter that describes the quality of measurements. This is because uncertainty fundamentally affects the decisions that are based upon the measurement result. Substantial progress has been made in devising procedures to estimate the uncertainty that originates in the analytical portion of the measurement, and guidance on these procedures is available [1]. However, a measurement almost invariably involves the process of taking a **sample**. This is because it is usually impossible to analyse the entire bulk of the material to be characterised (the **sampling target**). If the objective of the measurement is to estimate the value of the analyte concentration in a sampling target, then the uncertainty associated with the sampling process must inevitably contribute to the uncertainty associated with the reported result. It has become increasingly apparent that sampling is often the more important contribution to uncertainty and requires equally careful management and control. The uncertainty arising from the sampling process must therefore be evaluated. While existing guidance identifies sampling as a possible contribution to the uncertainty in a result, procedures for estimating the resulting uncertainty are not well developed and further, specific, guidance is required.

Historically, measurement scientists have been primarily concerned with measurements made within laboratories, and the process of sampling has been conducted by, and the responsibility of, a different set of people who are often in separate organisations. The measurement scientist's knowledge of the sampling process is then very limited. Conversely, the advent of *in situ* analytical techniques sometimes enables the measurement scientist to make measurements at the sampling site and in contact with the material to be sampled. Examples of this situation are process analysis within industrial production, and *in situ* measurements on contaminated land. The placing of the analytical sensor in these situations then constitutes the taking of a sample, and the measurement scientist becomes not only aware of, but responsible for, all stages of the measurement process, including the sampling. Such an awareness of the whole process is important, irrespective of the division of effort. Since analytical and sampling processes contribute to the uncertainty in the result, the uncertainty can only be estimated if there is an understanding of the complete process. Further, optimisation of the relative effort in sampling and analysis is only possible where sampling and analytical processes are both understood.

If the different stages are the responsibility of different people, there needs to be good communication between all of the parties involved. Sampling planners and analytical scientists need to optimise the whole measurement procedure, and to devise a strategy to estimate the uncertainty. Both need to discuss the objectives of the measurements with the customer. All three parties need guidance from the appropriate regulator on how these estimates of uncertainty are to be acted upon, to ensure the reliability of the decisions based upon the measurements. To underpin these decisions, all the parties need reliable estimates of uncertainty, including that arising from sampling. Although no general guidance can replace expert advice in complex or critical cases, this Guide describes the methods needed to fulfil the need for reliable estimates of uncertainty from sampling for most analytical measurement systems.

Summary

This Guide aims to describe various methods that can be used to estimate the uncertainty of measurement, particularly that arising from the processes of sampling and the physical preparation of samples. It takes a holistic view of the measurement process to include all of these steps as well as the analytical process, in the case where the measurand is defined in term of the value of the analyte concentration in the sampling target, rather than in just the sample delivered to the laboratory. The Guide begins by explaining the importance of knowing the total uncertainty in a measurement for making reliable interpretation of measurements, and judging their fitness for purpose. It covers the whole measurement process, defining each of the component steps, and describing the effects and errors that cause uncertainty in the final measurement.

Two main approaches to the estimation of uncertainty from sampling are described. The empirical approach uses repeated sampling and analysis, under various conditions, to quantify the effects caused by factors such as the heterogeneity of the analyte in the sampling target and variations in the application of one or more sampling protocols, to quantify uncertainty (and usually some of its component parts). The modelling approach uses a predefined model that identifies each of the component parts of the uncertainty, making estimates of each component, and sums them in order to make an overall estimate. Models from sampling theory can sometimes be used in this approach to estimate some of the components from a knowledge of the characteristics of particulate constituents.

Worked examples are given of each of these approaches, across a range of different application areas. These include investigations of the environment (of soil and water), of food (at growing and processing) and of animal feed. The estimates of the total uncertainty of measurement range from a few per cent up to 84% relative to the measurand. The contribution of the sampling is occasionally small but is often dominant (>90% of the total measurement variance). This suggests an increased proportion of the expenditure needs to be aimed at the sampling, rather than the chemical analysis, if the total uncertainty needs to be reduced in order to achieve fitness for purpose.

Management issues addressed include the responsibility of the quality of the whole measurement process, which needs to include the sampling procedure. Guidance is given on the selection of the most appropriate approach for any application, and whether one initial validation of the system is sufficient, or whether there is a need for ongoing monitoring of the uncertainty from sampling using quality control of sampling. The extra cost of estimating uncertainty is also considered in relation to the cost savings that can be made by knowing the uncertainty of measurement more reliably.

Such a Guide can never be fully comprehensive, and although there are appendices with details of some of the statistical techniques employed and sources of more detailed advice, there will often be a need for expert advice in more complex situations. This Guide aims to be a useful introduction to this subject, but we hope it will also stimulate further research into improved methods of uncertainty estimation.

Contents

List of acronyms and abbreviations

ANOVA	analysis of variance
AQC	analytical quality control
BIPM	Bureau International des Poids et Mesures
CEN	European Committee for Standardization
CH	constitution heterogeneity
CRM	certified reference material
CTS	collaborative trial in sampling
df	degrees of freedom
DH	distribution heterogeneity
FAPAS	trade name of body that organises international proficiency tests
FSE	fundamental sampling error
GEE	global estimation error
GFAAS	graphite furnace atomic absorption spectrometry
GSE	grouping and segregation error
GUM	Guide to the Expression of Uncertainty in Measurement. ISO
HPLC	high performance liquid chromatography
IDE	increment delimitation error
IEC	International Electrotechnical Commission
IFCC	International Federation of Clinical Chemistry and Laboratory Medicine
IPE	increment and sample preparation error
ISO	International Organization for Standardization
IUPAC	International Union of Pure and Applied Chemistry
IUPAP	International Union of Pure and Applied Physics
IXE	increment extraction error
LOD	limit of detection
LOQ	limit of quantification
MU	measurement uncertainty
NIFES	National Institute of Nutrition and Seafood Research
NIST	National Institute of Standards and Technology
OIML	International Organization of Legal Metrology
PSE	point selection error
PME	point mineralisation error
PT	proficiency test
QA	quality assurance
QC	quality control
RANOVA	robust analysis of variance
RSD	relative standard deviation
RST	reference sampling target
SD	standard deviation
SPT	sampling proficiency test
SS	sum of squares
SWE	weighting error
TAE	total analytical error
TSE	total sampling error

PART 1 – Introduction and scope

1 Introduction

1.1 Rationale for the Guide

The main purpose of measurement is to enable decisions to be made. The reliability of these decisions depends on knowing the **uncertainty** of the measurement results. If the uncertainty of measurements is underestimated, for example because the **sampling** is not taken into account, then erroneous decisions may be made that can have large financial consequences. The **fitness for purpose** of measurement results can only be judged by having reliable estimates of their uncertainty. For this reason it is essential that effective procedures are available for estimating the uncertainties arising from all parts of the measurement process. These must include uncertainties arising from any relevant sampling and physical preparation. Judgements on whether the analytical contribution to the uncertainty is acceptable can only be made with knowledge of the uncertainty originating in the rest of the measurement procedure.

1.2 Aim of the Guide

1.2.1 The aim of this Guide is to explain the rationale, and practical application, of the methods available for the estimation of uncertainty that includes the contribution from sampling. The Guide does not aim to recommend individual sampling protocols, which are often prescribed in other documents or regulations, but rather to consider the measurement uncertainty generated by whatever protocol is employed.

1.2.2 The Guide also aims to explain the importance of sampling to the overall uncertainty budget, and hence to the reliability of the consequent decisions made using the measurements. As well as explaining how to estimate the uncertainty, the Guide will explain the justification for including sampling in the overall management of the measurement process.

1.2.3 Unlike the assumption that is often made for estimates of uncertainty for an analytical method, an estimate for one sampling protocol for one batch of material should not be assumed as automatically applicable to any subsequent batch of material. For example, depending on the sampling target, the degree of **heterogeneity** (i.e. inhomogeneity) may have changed substantially. There will be a need, therefore, for routine monitoring of key parameters of sampling quality to examine and update estimates of uncertainty for subsequent batches.

1.3 Application to judging fitness for purpose

One of the main benefits of knowing the uncertainty of a measurement is to enable a stakeholder to judge its fitness for any particular purpose. A proper understanding of **uncertainty from sampling** must therefore be embedded in the broader perspective of fitness for purpose. This is important for two reasons. Firstly, it ensures that the estimate of uncertainty of each measurement is realistic when compared with the optimal value of uncertainty required to give reliable decisions. Secondly, given this level of uncertainty that is required to be fit for purpose, it is necessary to distribute effort (or expenditure) between the sampling and the analytical aspects of the measurement process in order to obtain the required uncertainty most economically. These ideas are developed further, and a quantitative approach to judging fitness for purpose by balancing uncertainty against cost is introduced, in Section 16.

1.4 Intended audience for the Guide

This Guide is intended primarily for specialists such as sampling planners and for analytical chemists who need to estimate the uncertainty associated with their measurement results. Other stakeholders should seek specialist advice for particular applications.

1.5 Relationship of this Guide to other documents

1.5.1 Current practice in the estimation of uncertainty for a broad range of measurements follows the 'Guide to the Expression of Uncertainty in Measurement' ('the GUM') [2], published in 1993 by ISO in collaboration with BIPM, IEC, IFCC, IUPAC, IUPAP and OIML. The GUM sets out the concepts required, established the general principles, and provided a procedure applicable to cases where an adequate model of the measurement process is available. The application of this approach to chemical analysis was described in 1995 in a Eurachem Guide for 'Quantifying Uncertainty in Analytical Measurement' [3], and broadened to include the use of validation and method performance data in a second edition in 2000 [1]. Other useful contributions to the practical estimation of uncertainty of analytical measurements using collaborative study data have been made by the Analytical Methods Committee of the Royal Society of Chemistry in 1995 [4], and by ISO TC/69 in 2004 [5]. This Guide on sampling is consistent with the general principles established in the GUM.

1.5.2 Sampling theory has developed largely independently of analytical chemistry and chemical metrology. Sampling quality has generally been addressed in sampling theory by the selection of a 'correct' sampling protocol, appropriate validation, and training of sampling personnel (i.e. **sampler**s) to ensure that this protocol is applied correctly [6]. It is then assumed that the samples will be representative and unbiased, and the variance will be that predicted by the model.

1.5.3 An alternative approach is to estimate the uncertainty of sampling for typical materials, or for **sampling target**s, during validation of the sampling protocol, and to confirm compliance in practice using ongoing quality control. This is more consistent with procedures already in place for the rest of the measurement process. Interestingly, the quality of sampling is only quantifiable through the measurements that are made upon the resultant samples.

1.5.4 Sampling protocols have been written to describe the recommended procedure for the sampling of innumerable types of material and for many different chemical components. These protocols are sometimes specified in regulation or in international agreements[a][7]. These procedures rarely identify the relative contributions of sampling and chemical analysis to the combined uncertainty.[b]

1.5.5 There is accordingly a range of prior literature on the theory and practice of sampling. As explained in Section 1.2.1, this Guide therefore does not seek to propose further sampling protocols but rather to provide methodologies to quantify the uncertainty that arises when a given protocol is used.

1.5.6 A handbook describing procedures for the estimation of uncertainty from sampling, derived from this Guide but with further case studies, has been prepared by the Nordtest group [8].

---

[a] The 'acceptance **sampling procedure**s' are applied to the sampling of a wide range of materials [7].

[b] Some concepts from sampling theory can be usefully adapted for estimation of uncertainty of measurement (Section 10.2).

1.6 Using the Guide

1.6.1 This document summarises the concepts necessary for understanding uncertainty in sampling, and provides procedures that allow their practical implementation. The Guide additionally covers issues related to management, quality assurance and reporting of results with uncertainty. The scope and intended field of application are set out in Section 2, which also summarises the approaches covered. Terminology is discussed in Section 3, and key terms defined in Appendix B.

1.6.2 Fundamental concepts are covered in Sections 4 and 5. An overview of the measurement process is provided in Section 4. This includes an explanation of the sampling terminology used, and indicates which steps in the process are considered in detail in this Guide. Measurement uncertainty and its sources are discussed further in Section 5.

1.6.3 Sections 6 to 10 describe methodologies for the estimation of uncertainty, with a discussion of the merits of the various options. The two broad approaches available are summarised in Section 6, and covered in detail in Sections 1 and 10 respectively. The intent is to provide a range of options that may be applied, rather than to specify any particular approach.

1.6.4 Management and quality issues are addressed in Sections 11 to 1. These include a very brief discussion of responsibilities for quality in sampling (Section 11) before discussing the selection of uncertainty estimation approach in Section 1. The use of sampling quality control to monitor sampling performance is covered in Section 1. Reporting and use of uncertainty, and its effect on the reliability of decisions, are discussed in Section 14. Cost is an important factor and selection of the most cost-effective and appropriate method of estimation is explained in Section 15. Knowing the value of the uncertainty helps to judge the fitness for purpose of the measurement as a whole, and its component parts, and this is discussed in Section 16.

1.6.5 A range of examples, a detailed glossary of terms and definitions used in this Guide, some important statistical procedures and experimental designs, and a discussion of improving sampling uncertainty using predictions from sampling theory are provided as appendices.

2 Scope and field of application

2.1 The principles of this Guide are applicable to the estimation of uncertainty from the full range of materials that are subject to analytical measurement (e.g. gaseous, liquid and solid). These include environmental media (e.g. rock, soil, water, air, waste and biota), foods, industrial materials (e.g. raw materials, process intermediaries and products), forensic materials and pharmaceuticals. This approach is applicable to sampling by any protocol, whether it uses single or composite samples, or single or multiple determinations.

2.2 The Guide describes the estimation of uncertainty using i) replicated measurement and sampling (the ‘empirical approach’) and ii) modelling based on identified influence quantities and theoretical considerations (the ‘modelling approach’).

2.3 The use of uncertainty estimates in the assessment of fitness for purpose and in the optimisation of effort among individual parts of the measurement process is covered. Methods of assessing fitness for purpose that are described include those based upon percentage of total variance and others based on cost-benefit analysis.

2.4 Although the general principles of this Guide apply, it does not specifically discuss microbiological sampling. Nor does it discuss the estimation of uncertainty in spatial or temporal information such as the location or size of areas of high analyte concentration.

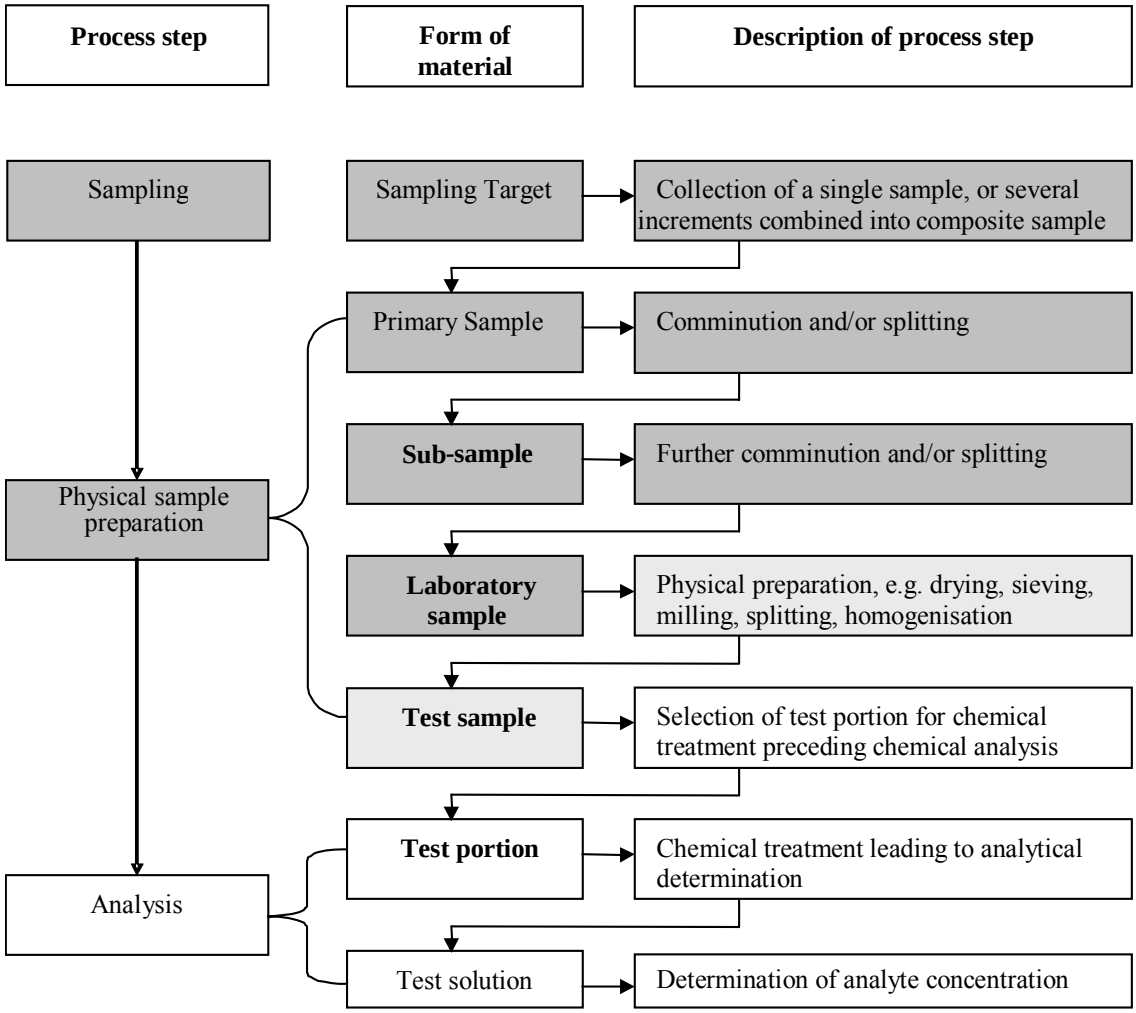
3 Terminology

3.1 The precise definitions of many of the terms used in this Guide vary depending on the area of application. A full listing of terms and their different definitions is given in Appendix B. In this Guide, normative definitions of each term have been selected that are as generally applicable as possible to all sectors of application. These terms are listed in Appendix B and in bold on first use in the text.

PART 2 – Fundamental concepts

4 Sampling in the measurement process

Figure 1: Schematic diagram of the typical measurement process



The figure shows a complete measurement process, starting with primary sampling and ending in the analytical determination. There are many intermediary steps, such as transportation and preservation of samples, not all of which are always present. Each step gives rise to a contribution towards the uncertainty of measurement. This Guide concentrates on the process steps of sampling and physical **sample preparation** (shaded boxes), as the last step is well covered in previous guidance [1]. Notice that two of the sampling steps occur within the laboratory (light grey) and are frequently considered to be part of the analytical process. For definitions of terms see Appendix B.

A sampling target is the portion of material, at a particular time, that the sample (and therefore the measurement result) is intended to represent. The sampling target needs to be defined prior to the design of the sampling plan. It may be defined by regulation, such as the whole of a batch, lot or consignment. If the properties and characteristics (e.g. analyte concentration) of the material in a certain area, or time period, are of interest and must be known, then it can be considered a sampling target. When the composition of a whole batch is required (e.g. of a food material), then the whole batch constitutes the target. When the spatial (or temporal)

variation of concentration is required (e.g. in finding 'hot spots' within a contaminated material), then each location where the concentration is required will be a separate sampling target. Any one sampling target will ultimately generate one reported measurement result and an uncertainty.

Primary samples are often made up of a number of **increment**s, which are combined to form a **composite sample** before a measurement is made. It is the uncertainty on this single measurement value, made on this composite sample, and caused by all of the preparatory steps, that is required. The value of this uncertainty will often be affected by the number of increments that are taken. This contrasts with the situation when several distinct primary samples (n) are taken from different parts of the sampling target, and measured separately. If the composition of the sampling target is calculated by taking the mean value of these separate measurements, then the uncertainty on the mean value is calculated using the standard error of the mean ($s/\sqrt{n}$). This is not the same as the uncertainty on the single measurement, the estimation of which is the objective of this Guide.

4.1 The whole process of measurement (Figure 1) typically begins with the taking of the primary sample from a sampling target. The resulting sample goes through one or more of a series of steps prior to the analytical determination. All steps contribute to uncertainty in the final result, when the analyte value required (i.e. **measurand** value or true value), is expressed in terms of the analyte concentration in the sampling target. Guidance already exists on the estimation of the analytical steps of the measurement process [1]. This will certainly include the selection of the test portion, the chemical treatment preceding measurement and the analytical determination, but may also include the physical preparation of the laboratory sample by means such as drying, sieving, milling, splitting and homogenisation.

4.2 In common practice, all the various portions of material in the second column of Figure 1 are often referred to simply as a 'sample'. It is clearly important to differentiate them carefully in discussion, especially those considered particularly in this Guide (in shaded boxes on Figure 1). This is discussed in more detail in Section 5.2.

4.3 Methods described in the Guide will help to identify the dominant source of the uncertainty, such as the sampling rather than the chemical analysis, but will not necessarily explain the cause. However, heterogeneity within the sampling target, either spatial or temporal, is known to be a significant cause of uncertainty in many circumstances. Separate studies would be needed to characterise the variability that contributes to the uncertainty. For the purpose of this Guide, heterogeneity within the sampling target is treated as just one cause of uncertainty in the final measurement. This is the case, whatever actions are taken to minimise the effects of the heterogeneity by the application of any particular sampling protocol.

5 Uncertainty of measurement

5.1 Definition of uncertainty of measurement

5.1.1 Uncertainty of measurement, or measurement uncertainty (MU), is defined in metrological terminology [2] as:

> *Parameter, associated with the result of a measurement, that characterises the dispersion of the values that could reasonably be attributed to the measurand.*

The definition includes several important features, which are discussed in the following paragraphs.

5.1.2 The 'parameter' may be, for example, a range, a standard deviation, an interval (like a confidence interval) or half-interval ($\pm u$ is a statement of a half-interval) or other measure of dispersion such as a relative standard deviation. Note that when MU is expressed as a standard deviation, the parameter is known as 'standard uncertainty', usually given the symbol u. Other forms of expression are considered in Section 14.

5.1.3 Uncertainty is 'associated with' each measurement result. A complete measurement result typically includes an indication of the uncertainty in the form $x \pm U$, where x is the measurement result and U an indication of the uncertainty (it will be seen that the symbol U has a special meaning, in this Guide; it indicates an 'expanded uncertainty', which will be discussed further in Section 14). This form of expressing a result is an indication to the end-user of the result that, with reasonable confidence, the result implies that the value of the measurand is within this interval.

5.1.4 The measurand is simply a quantity, such as a length, mass, or concentration of a material, which is being measured. The term 'value of the measurand' is closely related to the traditional concept of 'true value' in classical statistical terminology. From this alternative viewpoint 'uncertainty' has also been defined [9] as:

> *An estimate attached to a test result which characterises the range of values within which the true value is asserted to lie*

This definition (which will be referred to as the statistical definition) has the advantage of being easier to explain to decision makers, who often recognise the phrase 'true value' as the value of interest for their decision. It has the disadvantage that the true value itself can never be known and this generally requires further explanation.

5.1.5 The metrological definition asserts that uncertainty expresses 'the dispersion of the values that could reasonably be attributed to the measurand'. This is a particularly important phrase. It indicates that although the uncertainty is associated with a measurement result, the range quoted must relate to the possible range of values for the measurand. For example, the measurand could be the total mass of gold in a geological deposit. This is quite different from a statement of **precision**, which would describe the range of results that might be observed if the measurement were repeated. In requesting information about 'where the measurand value might be', this definition of uncertainty implicitly requires the measurement scientist to consider *all* the effects that might influence the measurement result. These effects obviously include the causes of random variation from one measurement to the next over the measurement timescale. But it is also essential to consider sources of **bias** during the experiment, and very often, these generate larger effects than can be observed by repeated measurement alone. That is, measurement uncertainty automatically asks for a range that includes an allowance for *both* random *and* systematic effects.

5.1.6 To consider a simple analytical example, a simple measurement of concentration in a solid will typically involve extraction of material, weighings, volumetric operations and perhaps spectrometry or chromatography. Repeated measurement will show a spread of values due to random variations in these operations. But all analysts know that extraction is rarely complete, and, for a given material, that failure to extract material will lead to a consistently low result. While good analytical practice always attempts to reduce such effects to insignificance, some bias will remain. In expressing the uncertainty about the value of the measurand, then, the analyst must take into account the reasonable possibility of bias from

such causes. (Usually, this is done by considering such information as the range of analyte recoveries observed on reference materials or from spiking experiments.)

5.1.7 The same considerations apply in the case of sampling. It is well known that different samples taken from a bulk material will often show real variation in value, which is clear from repeated measurement. It is also well known that sampling may be biased, for example by differential removal of materials, inappropriate timing of sampling where temporal fluctuations occur, or by access restrictions. These effects will influence the relationship between the value of the measurand and the result that is observed. While good practice in sampling is intended to reduce these effects to insignificance, a careful assessment of uncertainty always considers the possibility of residual systematic effects.

5.1.8 Current guidance on measurement uncertainty [2] makes it clear that uncertainty of measurement (Section 2.2 of reference [2]) is not intended to allow for 'gross error'. This would preclude, for example, mistakes caused by transcription errors or gross misuses of the measurement protocol. Sampling can, however, produce high levels of uncertainty (e.g. 80% of the concentration value), simply through the routine application of an accepted measurement protocol to a highly heterogeneous material. Even when procedures are nominally correct, there will also be slight variations in the actual procedures due to the ambiguity in the measurement protocols, and the minor adaptations that are made to protocols in real-world sampling situations. Whether these high levels of uncertainty lead to unacceptable levels of reliability in the decisions that are based upon them, depends upon a rigorous evaluation of fitness for purpose (see Section 16).

5.2 Specification of measurand

5.2.1 When an end-user is presented with a concentration result quoted for a bulk sample in the form '$x \pm U$', they will very naturally interpret that interval as including the range of values attributable to the concentration in the sampling target (e.g. a batch of material). Implicit in this view is the idea that the measurand is 'the (true) concentration (of the analyte) in the batch of material', and that the uncertainty includes any necessary allowance for heterogeneity in the bulk. The analyst, by contrast, might refer to 'the concentration in the laboratory sample analysed', implicitly ruling out the variation between laboratory samples. Clearly, one viewpoint includes the effects of sampling, while the other does not. The effect on the uncertainty can, of course, be very considerable. In metrological terms, this distinction arises because the two views are considering different measurands. One is considering 'concentration in the sampling target', the other 'concentration in the laboratory sample'. Another example might be 'contaminant concentration at a factory outlet at the time of sampling', compared to 'the average contaminant concentration over a year'.

5.2.2 These ambiguities in interpretation can be avoided only by careful specification of the measurand. It is clearly necessary to state the quantity (mass, length, concentration etc.). It is equally important to be clear on the scope of the measurement, by including information on factors such as the time, location, or population to which the measurement result will be assumed to apply. Some particular instances of measurand specification and their implications for uncertainty estimation are discussed below.

It is never possible to avoid all ambiguity in implementing the wording of the sampling protocol.

5.2.3 When a composite sample is taken by the combination of several increments from across a sampling target, and analysed as a single primary sample, that single determination of analyte concentration provides an estimate of the value of the measurand (i.e. the average composition of the target), as discussed briefly in Section 4. The uncertainty on this single

value reflects the uncertainty in the estimate of the measurand value. In contrast, if several independent primary samples are taken from the target, each analysed once, and the mean value calculated, this mean value will also be an estimate of the value of the measurand. However, the uncertainty will not be that of the measurement (expressed as standard deviation, s), but the standard error of the mean value (expressed as $s/\sqrt{n}$). This later uncertainty on the mean can be reduced by taking more primary samples,[c] whereas the uncertainty on the measurement cannot.

5.3 Error, precision and uncertainty

5.3.1 Uncertainty is related to other concepts, such as accuracy, error, trueness, bias and precision. Other guidance discusses the relationships in some detail [1, 2]. However, it is worth repeating some of the important differences:

- Uncertainty is a range of values attributable on the basis of the measurement result and other known effects, whereas error is a single difference between a result and a 'true (or reference) value'
- Uncertainty includes allowances for all effects that may influence a result (i.e. both random and systematic errors); precision only includes the effects that vary during the observations (i.e. only some random errors).
- Uncertainty is valid for correct application of measurement and sampling procedures, but, as noted in Section 5.1.8, it is not intended to make allowance for gross operator error

5.4 Sampling and physical preparation as sources of uncertainty of measurement

5.4.1 The act of taking a sample introduces uncertainty into the reported measurement result wherever the objective of the measurement is defined in terms of the analyte concentration in the sampling target and not simply in the laboratory sample.

5.4.2 Sampling protocols are never perfect in that they can never describe the action required by the sampler for every possible eventuality that may arise in the real world in which sampling occurs. The location in space (or time) for the taking of a sample is rarely specified exactly (e.g. to the nearest millimetre or second). The sampler has to make such decisions (ideally on objective criteria), but as heterogeneity is inevitable (in space or time) such decisions will affect the estimated concentration. An appreciation of these sources of uncertainty is important in the design and implementation of methods for the estimation of uncertainty. When **duplicate sample**s are taken, for example, taking them at exactly the *same* place and time may not reflect the uncertainty of the measurement that really exists. This will be discussed further in the description of methods of estimation (Sections 6 to 10), and in the various worked examples (Appendix A).

5.4.3 Heterogeneity always gives rise to uncertainty. If the sampling target were perfectly homogeneous then this contribution would be zero, but nearly all materials are heterogeneous to some extent at some scale. If the test portion is a few micrograms, then nearly all material will be heterogeneous and the sampling step will contribute to the uncertainty in the measurement of an analyte concentration. Heterogeneity can be quantified in a separate experiment, but if the aim is to estimate the analyte concentration in the larger sampling target, then this heterogeneity is just one cause of measurement uncertainty (as discussed in Section 4.2).

[c] Assuming that the samples are random and independent, and assuming zero bias.

5.4.4 Similar arguments can be made for the uncertainty that arises in the processes of physical preparation (e.g. transportation, preservation, comminution, splitting, drying, sieving, homogenisation) that happen after the act of sampling and before the chemical preparation of the test sample (Figure 1). Each step can introduce errors from a range of mechanisms, such as loss of analyte, loss of fine particles, or contamination from equipment or previous samples. The methods employed, and training given, should aim to reduce these errors to a minimum. In addition, however, procedures are required to estimate the uncertainty that all of these steps, when applied in practice, generate in the final measurement value.

5.5 Sources of uncertainty

5.5.1 Uncertainty arises from a variety of sources, and these have been categorised in different ways. For example, the Eurachem Uncertainty Guide identifies eight major categories of effects that are important in estimating uncertainty have been identified [3], of which the first two are sampling and sample preparation. Specific effects identifiable in these two categories are shown in Table 1. A modelling approach might use these effects as the basis for a mathematical model. Alternatively, sampling theory identifies eight distinct sources of error in sampling (Table 2); each of these can also be reduced to a variety of causal factors, which in turn can be used in various modelling approaches. A further alternative approach is to consider all of the steps in the measurement process (Figure 1) as sources of uncertainty that make some contribution to the uncertainty of the final measurement. In this Guide, the simplest study designs treat uncertainty as arising from four classes of effect (Table 3), and the classes are treated as sources of uncertainty in a simple statistical model; this is consistent with the grouping of uncertainty sources explicitly suggested in reference [3]. In its simplest form, this categorisation can be reduced to two categories: '**sampling uncertainty**' and 'analytical uncertainty'.

5.5.2 The important feature of each of these different classifications is that each is intended to ensure that, however they are grouped and evaluated, all practically important effects are taken into account in estimating the uncertainty. As long as this requirement is met, any categorisation scheme may be applied to the estimation of uncertainty. The categorisation schemes listed in Table 2 and Table 3 cover all practically important effects.

5.5.3 Each different categorisation of sources will generally lead to a different study design, and very often to fundamentally different methods of evaluation of uncertainty contributions. This results in substantially independent estimates of uncertainty via different approaches. As noted elsewhere [5], grossly different estimates of uncertainty for the same system suggest that at least one study methodology is in error. This forms the basis of a check on the validity of an approach. Where practicable, therefore, comparison of uncertainty estimates arising from independent evaluation approaches is recommended as a means of validating particular estimates and of assessing the validity of different approaches.

5.6 Heterogeneity as a source of uncertainty

5.6.1 IUPAC currently define both homogeneity and heterogeneity as 'The degree to which a property or constituent is uniformly distributed throughout a quantity of material.' (see Appendix B for definitions). So defined, heterogeneity is among the most important factors contributing to uncertainty associated with sampling. Increments from different locations in the sampling target will have different concentrations of analyte in a heterogeneous material and there will be a sample-to-sample variation in analyte concentration – usually visible as a contribution to the observed variation of results. In general, the exact dependence of concentration on location is unknown, so no correction can

be made. This results in uncertainty in any given result or, in general, any average of such results.

5.6.2 IUPAC note, as an addendum to the above definition, that 'The degree of heterogeneity (the opposite of homogeneity) is the determining factor of sampling error.' The note is a good indication of the importance of heterogeneity in sampling. There are other sources of error and uncertainty in the general operation of sampling; for example, cross-contamination and imperfect stabilisation of samples, either of which can result in (unknown) bias or additional variability. Yet heterogeneity and its effects – such as random variability and selection bias – remain the largest problem in properly managed sampling and will generally be the most significant source of uncertainty.

5.6.3 An alternative definition of homogeneity is sometimes used for particulate material, which, if it consists of particles of different materials, cannot ever be 'homogeneous' in the sense defined by IUPAC. In this context, a mixture in which the probability of selection of different types of particle is constant throughout the sampling target may be termed 'homogeneous' to denote that the expected concentration in would be the same in a sample taken at any point in the material. Even here, however, it must be recognised that the particulate nature of the material leads to sample-to-sample variation due to slightly different composition of the increments actually taken; heterogeneity, as defined by IUPAC, still has an effect under these circumstances, and consequently still contributes to the uncertainty.

Table 1: Some sources of uncertainty in sampling and sample preparation, adapted from reference [3]

Sampling	**Sample preparation**
- Heterogeneity (or inhomogeneity) - Effects of specific sampling strategy (e.g. random, stratified random, proportional etc.) - Effects of movement of bulk medium (particularly density selection) - Physical state of bulk (solid, liquid, gas) - Temperature and pressure effects - Effects of sampling process on composition (e.g. differential adsorption in sampling system) - Transportation and preservation of sample	- Homogenisation and/or **sub-sampling** effects - Drying - Milling - Dissolution - Extraction - Contamination - Derivatisation (chemical effects) - Dilution errors - (Pre-)Concentration - Control of speciation effects

Table 2: Sources of sampling uncertainty in sampling theory*

Source	Description
Fundamental sampling error (FSE)	A result of the constitutional heterogeneity (the particles being chemically or physically different)
Grouping and segregation error (GSE)	A result of the distributional heterogeneity
Long-range point selection error (PSE_1)	Trends across space or over time
Periodic point selection error (PSE_2)	Periodic levels across space or over time
Increment delimitation error (IDE)	Identifying the correct sample to take. Considers the volume boundaries of a correct sampling device
Increment extraction error (IXE)	Removing the intended sample. Considers the shape of the sampling device cutting edges
Increment and sample preparation error (IPE)	*Contamination* (extraneous material in sample): *Losses* (adsorption, condensation, precipitation etc.): *Alteration of chemical composition* (preservation): *Alteration of physical composition* (agglomeration, breaking of particles, moisture etc.): ***Involuntary mistakes* (mixed sample numbers, lack of knowledge, negligence): ***Deliberate faults* (salting of gold ores, deliberate errors in increment delimitation, forgery etc.)
Weighting error (SWE)	The result of errors in assigning weights to different parts of an unequal composite sample

*This classification follows that of Gy [17] and others (discussed further in Section10).
** Excluded from uncertainty estimates as gross errors [2].

Table 3: Uncertainty contributions in the empirical approach

Process	Effect class*	
	***Random** (precision)*	***Systematic** (bias)*
Analysis	Analytical variability (combined contribution of random effects)	Analytical bias (combined effect of bias sources)
Sampling	Sampling variability (dominated by heterogeneity and operator variations)	**Sampling bias** (combined effect of selection bias, operator bias etc.)

*The differentiation of random from systematic effects can depend on the context. A systematic effect in measurements by one organisation (e.g. analytical bias) can also be considered a random effect when viewed in the context of the consensus value from an inter-organisational proficiency test.

PART 3 – Estimation of measurement uncertainty including sampling

6 Approaches to uncertainty estimation

6.1 There are two broad approaches to the estimation of uncertainty. One of them, described as 'empirical', 'experimental', 'retrospective', or 'top-down', uses some level of replication of the whole measurement procedure to give a direct estimate of the uncertainty for the final result of the measurement. This approach is called the 'empirical' approach in this Guide. The second, variously described as 'modelling', 'theoretical', 'predictive' or 'bottom-up', aims to quantify all of the sources of uncertainty individually, and then uses a model to combine them. It will accordingly be referred to as the 'modelling' approach. These approaches are not mutually exclusive. The empirical method can be adapted to estimate contributions to uncertainty from one or more effects or classes of effect. Both approaches can usefully be used together to study the same measurement system, if required. The applicability of the two approaches varies between the different materials to be sampled.

6.2 The approach taken in this Guide is to describe in detail the empirical approach, which has the widest applicability to the broadest range of measurement systems and applications (e.g. gaseous, liquid and solid). Modelling approaches are described for particular situations to which they apply (e.g. particulate solids). Advice will also be given on how a combination of these different approaches can be used to give more reliable and cost-effective estimates of uncertainty in a range of measurement systems. This dual approach is intended to enable a user of the Guide to select the most appropriate method of uncertainty estimation for their particular situation. (Section 1 provides guidance on selection of the approach.)

6.3 Reference [5] notes that the modelling approaches and the type of empirical study used in collaborative trial are extremes of a continuum:

> *Note, however, that observed repeatability or some other precision estimate is very often taken as a separate contribution to uncertainty even in the [modelling] approach. Similarly, individual effects are usually at least checked for significance or quantified prior to assessing reproducibility. Practical uncertainty estimates accordingly often use some elements of both extremes.*

In referring to either extreme, therefore, it is important to be aware that these are extremes and that many practical estimates involve elements of both approaches.

6.4 The overall objective of any approach is to obtain a sufficiently reliable estimate of the overall uncertainty of measurement. This need not necessarily require all of the individual sources of uncertainty to be quantified, only that the combined effect be assessed. If, however, the overall level of uncertainty is found to be unacceptable (i.e. the measurements are not fit for purpose) then action must be taken to reduce the uncertainty. Alternatively, the uncertainty may be unnecessarily small, in which case there may be justification for increasing the analytical uncertainty, and thereby decreasing the cost of analysis. Methods for modifying uncertainty are discussed in Appendix E. At this stage, however, it is essential to have information on which general part of the measurement procedure is causing the dominant contribution to the overall uncertainty, and it may then be necessary to evaluate individual effects. The advantage of detailed early study is that this information is already available; the disadvantage is that it is costly to obtain and may prove unnecessary if uncertainty is acceptable. Planners should accordingly consider the level of detail required in

an uncertainty estimate, taking account of the probability of requiring detailed information for further development.

7 The measurand

7.1 In the following discussion, it is assumed that the measurand is an average value representing the composition of the whole sampling target, and that the measurand is to be estimated through a process of sampling and analysis. This relates to the specification of the measurand (Section 5.2) and the definition of the sampling target (Section 4.1).

8 General conduct of studies of sampling uncertainty

8.1 Analytical work should be undertaken under an appropriate quality system, including validated analytical methods, proficiency testing, internal quality control and external assessment where appropriate. Validation procedures should include all the steps normally undertaken within the laboratory (including any sub-sampling of test samples), and should include checks on bias using certified reference materials, or other methods, for the estimation of analytical bias. Note that the uncertainty estimation methods described in this Guide can also be applied to the estimation of uncertainties associated with sub-sampling.

8.2 Laboratories undertaking the chemical analysis should report the concentration estimates exactly as found; in particular, values must not be censored, truncated or reported as 'less than' a reporting limit, whether below the limit of detection (LOD) or below zero. Failing to report negative or sub-LOD observations will result in an underestimate of the uncertainty.

9 Empirical approach

9.1 Overview

9.1.1 The empirical ('top-down') approach is intended to obtain a reliable estimate of the uncertainty, without necessarily knowing any of the sources individually. It relies on overall reproducibility estimates from either in-house or inter-organisational measurement trials. It is possible to describe the general type of source, such as random or systematic effects, and to subdivide these as those arising from the sampling process or the analytical process. Estimates of the magnitude of each of these effects can be made separately from the properties of the measurement methods, such as **sampling precision** (for random effects arising from sampling) or analytical bias (for systematic effects arising from chemical analysis). These estimates can be combined to produce an estimate of the uncertainty in the measurement result. This approach is illustrated in detail in Examples A1, A2, A3 and A4.

9.2 Uncertainty sources

9.2.1 It is possible to consider uncertainty of measurements to arise from four broad sources of error. These four sources are the random errors arising from the methods of both the sampling and analysis, and also the systematic errors arising from these methods. These errors have traditionally been quantified as the sampling precision, analytical precision, sampling bias and the analytical bias respectively (Table 4). If errors from these four sources are quantified, separately or in combinations, it is possible to estimate the uncertainty of the measurements that these methods produce. Methods for the estimation of three of the four errors are well established. Sampling and analytical precision can be estimated by duplication

of a proportion (e.g. 10%) of the samples and analyses respectively. Analytical bias can be estimated by measuring the bias on well-matched certified reference materials, and assuming that this bias represents that present for the test material, or by taking it directly from the validation of the analytical method.

9.2.2 Procedures for estimating sampling bias include the use of a **reference sampling target** [10, 24] (the sampling equivalent of a reference material). Alternatively they utilise measurements from inter-organisational sampling trials, in which the sampling bias potentially introduced by each participant is included in the estimate of uncertainty based on the overall variability (Section 9.5). Although some of the components of uncertainty associated with systematic effects may be difficult to estimate, it may be unnecessary to do so if there is good evidence that systematic effects are small and under good control. Such evidence may be qualitative, as in prior knowledge of the chemical or physical nature of the sampling target, or quantitative, such as information, for example from prior measurements on complete batches. (See Examples A3 and A4, Appendix A.)

Table 4: Estimation of uncertainty contributions in the empirical approach

Process	**Effect class**	
	Random *(precision)*	***Systematic*** *(bias)*
Analysis	e.g. *duplicate analyses*	e.g. *certified reference materials*
Sampling	*Duplicate samples*	*Reference sampling target, inter-organisational sampling trial*

Four classes of effects that contribute to the uncertainty of measurements, and methods for their estimation.

9.3 Statistical model for the empirical estimation of uncertainty

In order to design experimental methods to estimate uncertainty using the empirical approach it is necessary to have a statistical model describing the relationship between the measured and true values of analyte concentration. This random effects model considers a single measurement of analyte concentration (x), on one sample (composite or single), from one particular sampling target:

$$x = X_{true} + \varepsilon_{sampling} + \varepsilon_{analysis}$$

where X_{true} is the true value of the analyte concentration in the sampling target (i.e. equivalent to the value of the measurand). For example, this could be the total mass of the analyte in the target divided by the total mass of the target. The total error due to sampling is $\varepsilon_{sampling}$ and the total analytical error is $\varepsilon_{analysis}$

In an investigation of a single sampling target, if the sources of variation are independent, the measurement variance $\sigma 2_{meas}$ is given by,

$$\sigma^2{}_{meas} = \sigma^2{}_{sampling} + \sigma^2{}_{analytical}$$

where $\sigma^2{}_{sampling}$ is the between-sample variance on one target (largely due to analyte heterogeneity), and $\sigma^2{}_{analysis}$ is the between-analysis variance on one sample.

If statistical estimates of variance (s^2) are used to approximate these parameters, we get

$$s^2_{meas} = s^2_{sampling} + s^2_{analytical}$$

The standard uncertainty (u) can be estimated using s_{meas}, which is therefore given by

$$u = s_{meas} = \sqrt{s^2_{sampling} + s^2_{analytical}} \qquad \text{Equation 1}$$

Variance caused by physical sample preparation can be included into the sampling variance, or expressed as a separate term if required.

In a survey across several sampling targets, which is recommended for the estimation of sampling uncertainty (Section 9.4.2), the model needs to be extended to

$$x = X_{true} + \varepsilon_{target} + \varepsilon_{sampling} + \varepsilon_{analysis}$$

where the additional term ε_{target} represents the variation of concentration between the targets and has variance $\sigma^2_{between-target}$. Appropriate ANOVA generates estimates of the variances $\sigma^2_{between-target}$, $\sigma^2_{sampling}$ and $\sigma^2_{analysis}$, and the uncertainty is estimated exactly as before, using Equation 1.

The total variance σ^2_{total}, given by

$$\sigma^2_{total} = \sigma^2_{between-target} + \sigma^2_{sampling} + \sigma^2_{analytical}$$

is also a useful parameter in assessing fitness for purpose; this is discussed further in section 16.2. For practical purposes the population variances are replaced by their estimates s^2 to give

$$s^2_{total} = s^2_{between-target} + s^2_{sampling} + s^2_{analytical} \qquad \text{Equation 2}$$

9.4 Empirical estimation of uncertainty

9.4.1 Four types of method are applicable to the estimation of uncertainty using the empirical approach (Table 5). A fifth variographic method is described briefly in Section 9.6. The main method described further in this Guide is the 'duplicate method' (#1). If one sampler uses several sampling protocols in Method #2, any bias between the protocols can be detected. If multiple samplers all apply one protocol (Method #3, which is equivalent to a collaborative trial in sampling – CTS, or method performance test), then bias between different samplers can be detected and included in the uncertainty estimate. If multiple samplers apply different protocols that are selected to be the most appropriate for the stated objective, in their professional opinion (Method #4, which is equivalent to a sampling proficiency test – SPT), then any sampling bias introduced by either the sampling protocol, or the sampler, can be detected and included in the estimate of uncertainty.

Table 5: Four empirical methods for estimating combined uncertainty including sampling

Method #	Method description	Samplers (persons)	Protocols	Component estimated			
				Sampling		Analytical	
				Precision	Bias	Precision	Bias
1	Duplicates	Single	Single	Yes	No	Yes	No[1]
2	Protocols	Single	Multiple	Between protocols		Yes	No [1]
3	CTS	Multiple	Single	Between samplers		Yes	Yes [2]
4	SPT	Multiple	Multiple	Between protocols +between samplers		Yes	Yes [2]

[1]Analytical bias information may be obtained by including certified reference materials in the analytical run (see Example A2, Appendix A).

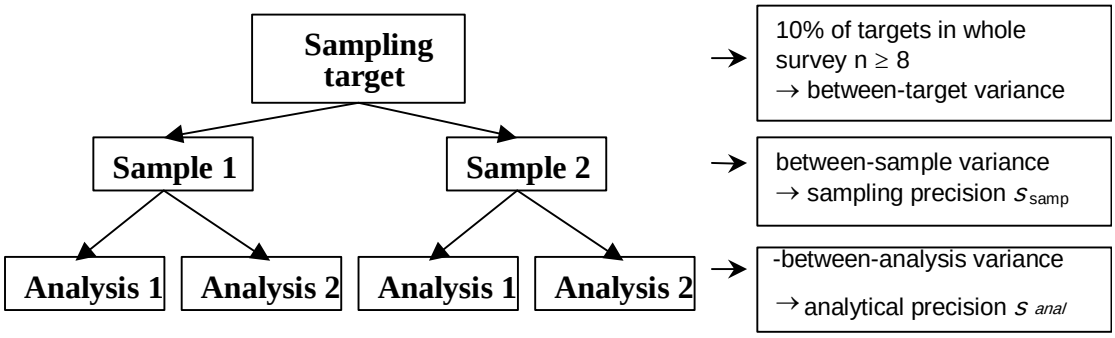
[2]Analytical bias is partially or completely included in collaborative exercises where multiple laboratories are involved.

9.4.2 The duplicate method is the simplest and probably most cost-effective of the four methods described in Table 5. It is based upon a single sampler duplicating a small proportion (i.e. 10%, but no less than eight targets) of the primary samples[d] [11, 12]. Ideally the duplicates are taken from at least eight sampling targets, selected at random to represent the typical composition of such targets. If only one target exists, then all eight duplicates can be taken from it, but the uncertainty estimate will only be applicable to that one target. The duplicated samples are taken by repeating the same nominal sampling protocol, with permitted variations that reflect the ambiguity in the sampling protocol and the effect of small-scale heterogeneity of the analyte of interest on the implementation of that protocol. For example, in a 'W' design for collecting a composite sample over a bay of lettuces, the initial starting point and orientation of the 'W' is altered for the duplicate sample; for a grid design, again, starting point and orientation are altered (ExampleA1, Appendix A). The duplicate samples are obtained using a single sampling protocol and by a single person (sampler). Both of the duplicated samples are subject to physical preparation resulting in two separate test samples. Duplicate test portions are drawn from both of the test samples and analysed in duplicate (i.e. duplicate chemical analysis). This system of duplicated sampling and chemical analysis is known as a 'balanced design' (Figure 2). Note that the duplicate method does not include any contribution from sampling bias, which must be either assumed to be negligible, or estimated separately using, for example, multiple samplers, multiple protocols and/or inter-organisational sampling trials as in the other three methods.

Note: Although the 'duplicate method' is generally described in terms of a single sampler and protocol, the same design can be used with different samplers to incorporate the 'between-operator' contribution to uncertainty (equivalent to Method #3).

[d] A higher level of replication can be used, but duplication is usually the most effective form of replication in sampling studies. It is better to take duplicates from 12 sample targets, than take triplicates from eight targets, for example, as although each estimate of the uncertainty of sampling ($s_{sampling}$) has a lower standard error, the estimate is based on a smaller proportion of the entire population of sample targets, and is therefore less representative. The minimum number of eight duplicates is required to provide sufficiently reliable estimates of the uncertainty [12].

Figure 2: A balanced design



Balanced experimental design for empirical estimation of uncertainty (i.e. two-stage nested design), using the 'duplicate method'.

9.4.3 The test portions are then chemically analysed anonymously by an appropriate analytical method under repeatability conditions (e.g. distributed randomly within an analytical batch). If estimates of the analytical portion of the measurement uncertainty have been made independently by the laboratory, this will be useful for comparison with estimates made by this method. Variance caused by physical sample preparation can be included into the sampling variance by having independent preparation on each of the sample duplicates. Alternatively this variance can be estimated separately by inserting an extra level of replication in the experimental design (Appendix D).

9.4.4 The balanced design proposed here will only give the repeatability standard deviation of the analytical measurements. In order to estimate the other part of the analytical uncertainty, an allowance has to be made for potential analytical bias. The limitations of this approach, and a worked example, are given in Section 6 of Example A2. One alternative is to ask the measuring laboratory for the repeatability and measurement uncertainty, and then to check that the repeatability obtained in this study is similar to that claimed by the laboratory. If this is the case, we can use the measurement uncertainty given by the lab as u(analytical) (normally $U/2$). A second alternative is to use the estimation of analytical bias made from the well-matched certified reference materials contained in the analytical batch. This bias estimate can then be combined with the repeatability to obtain the measurement uncertainty [1, 30].

9.5 Calculation of uncertainty and its components

9.5.1 The random component of the uncertainty can be estimated by applying analysis of variance (ANOVA)[e] or range calculation[f] to the measurements of concentration on the duplicated samples. The estimation is based upon the model described in Section 9.3, applied to whatever measurement protocol that is being employed (with its specified number of sample increments and analytical replicates).

[e] There is often a small proportion (i.e. <10%) of outlying values in the frequency distributions of the analytical, within-sample and between-sample variability. This requires the use of some method of down-weighting the effect of the outlying values on classical ANOVA, such as the use of robust statistical methods. This gives a more reliable estimate of the variances of the underlying populations. A fuller explanation of these methods is given in the worked example in Appendix A1 (p40), A2 (p45).

[f] See example in Appendix A2.

9.5.2 The values of s_{samp} and s_{anal} from the ANOVA are estimates of sampling precision and analytical precision respectively. The random component of the measurement uncertainty is calculated by the combination of these two estimates (Equation 1). The expanded uncertainty, for approximately 95% confidence for example, requires this value to be multiplied by a coverage factor of 2. The expanded uncertainty (U) is then calculated using

$$U = 2s_{meas} \qquad \text{Equation 3}$$

U can also be expressed relative to the reported value x and expressed in terms of a percentage, as a relative expanded uncertainty U':

$$U' = 100\frac{2s_{meas}}{x}\% \qquad \text{Equation 4}$$

The relative uncertainty is more widely applicable than the standard uncertainty, as it does not change appreciably as a function of concentration at values well above the analytical detection limit (>10 times). Other coverage factors can be selected as appropriate. The improvement of this estimate of uncertainty to include the systematic error from the chemical analysis is discussed in Example A2, Appendix A.

The relative expanded uncertainty for the sampling or analysis alone can similarly be expressed as

$$U_{samp}' = 100\frac{2s_{samp}}{x}\%$$

and

$$U_{anal}' = 100\frac{2s_{anal}}{x}\%$$

9.5.3 Because the uncertainty of many measurement systems is dominated by heterogeneity within the sampling target, the use of the simplest ‘duplicate method’ often gives a reasonably reliable estimate of uncertainty. Studies of environmental systems have shown that between-operator and between-protocol effects are often much smaller than those caused by heterogeneity [43]. Further information on selection of the most effective method for uncertainty estimation is provided in Section 1. Examples of applications of the duplicate method are given in Examples A1 and A2, Appendix A.

9.5.4 In addition to an initial single estimate of uncertainty for a particular sampling protocol applied to a particular sampling target, routine application of the ‘duplicate method’ is also useful as a way of monitoring the ongoing sampling quality (Section 13). This can allow for the effect on uncertainty of changes in the heterogeneity of the sampling target between different applications of the same sampling protocol. Quantitative evidence of the quality of sampling can then be gained, rather than relying solely on the assumption that samples are representative, if taken by a correct protocol.

9.6 Alternative empirical methods of uncertainty estimation

9.6.1 Variography has also been suggested as a further empirical means of estimating the uncertainty of measurement from the combined sources of sampling and analysis [13]. It is particularly useful in situations where there is large-scale spatial and/or temporal variation in contaminant concentration that can be quantified and modelled. This is the case for some instances of rock and soil geochemistry, and in emission control of (e.g. waste water), when

large numbers (n>100) of evenly distributed samples have been taken. Further guidance on the principles and application of variography for this purpose, with a case study, is available [8].

10 The modelling approach

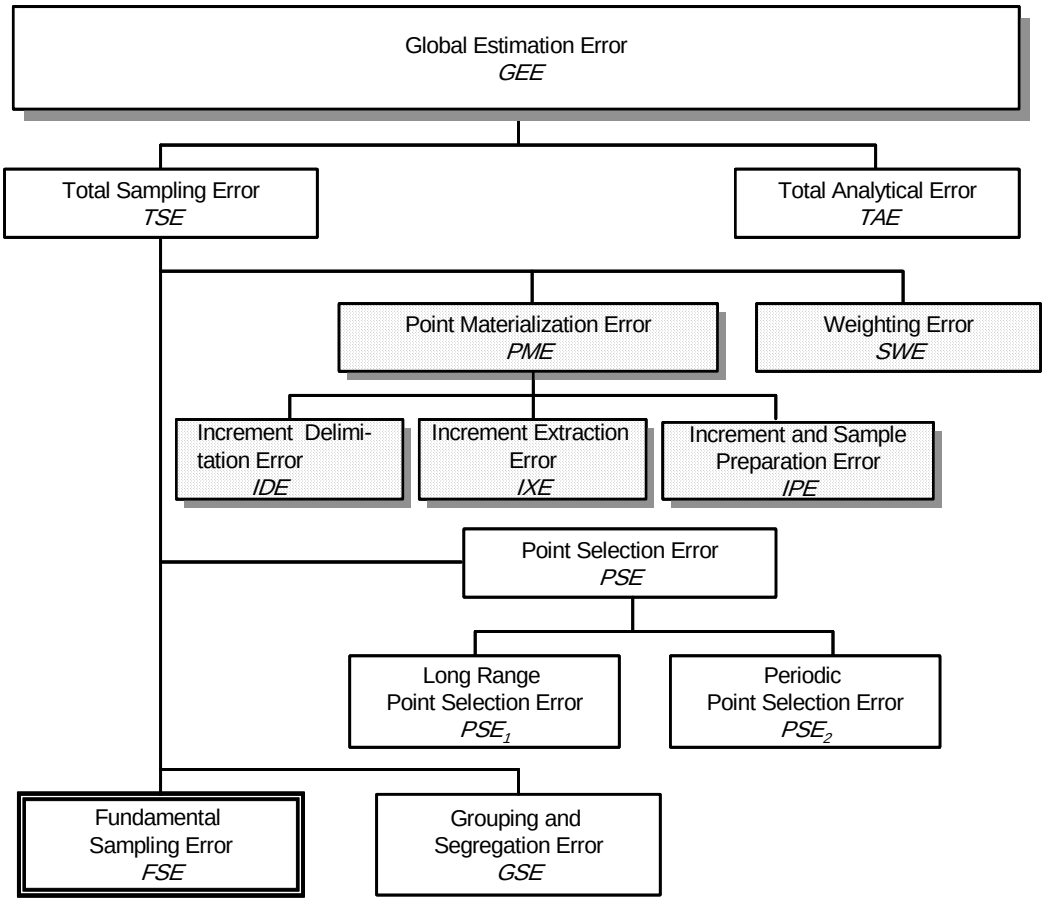
10.1 Cause-and-effect modelling

10.1.1 The modelling approach, often colloquially known as 'bottom-up', has been described for measurement methods in general [2], and applied to analytical measurements [1]. It initially identifies all of the sources of uncertainty, quantifies the contributions from each source, and then combines all of the contributions, as a budget, to give an estimate of the combined standard uncertainty. In the process, the measurement method is separated into all of its individual steps. This can usefully take the form of a cause-and-effect, or 'fish-bone', diagram [3]. The uncertainty of measurement generated by each of these steps is estimated independently, either empirically or by other methods. The combined uncertainty is then calculated by combining the uncertainty from all of the steps by appropriate methods. This approach is well established for analytical methods [1], but has only recently been applied to the process of sampling [13, 14]. For particulate systems, sampling theory uses a similar approach to identifying seven types a sampling error. One of these errors (fundamental) is estimated using an equation based on a detailed knowledge of the individual particles being sampled, as discussed in the next section (and Example A5, Appendix A).

10.2 Sampling theory for estimation of uncertainty

10.2.1 Sampling theory has been proposed as an appropriate method for estimating uncertainty from sampling [15]. This approach relies on the use of a theoretical model, such as that of Gy. Pierre Gy has developed a complete sampling theory described in many publications [6, 16, 17, 18, 19, 20], including its latest developments [19]. Figure 3 shows Gy's classification of sampling errors. Most sampling errors, except the preparation errors, are due to the material heterogeneity, which can be divided into two classes: 1) constitution heterogeneity (CH), and 2) distribution heterogeneity (DH). Both heterogeneities can be mathematically defined and experimentally estimated. Constitution heterogeneity refers to the fact that all natural materials are heterogeneous, that is, they consist of different types of particles (molecules, ions, grains). The distribution is heterogeneous if the particles are not randomly distributed in the sampling target (or lot) to be investigated.

Figure 3: Classification of sampling errors in Gy's sampling theory



$$GEE = TSE + TAE$$

$$TSE = (PSE + FSE + GSE) + (IDE + IXE + IPE) + SWE$$

* incorrect sampling errors are indicated by shaded boxes, and are excluded from the estimates of uncertainty

10.2.2 The classification of errors of sampling forms a logical and useful framework for designing and auditing sampling procedures. Those that are central to the estimation of uncertainty (e.g. FSE in Figure 3) are discussed below, and others (SWE, PSE and GSE) in Appendix C.

10.2.3 The total determination error, which Gy calls the global estimation error (GEE), is the sum of the total sampling error (TSE) and total analytical error (TAE). The components of TSE can be divided into two major groups: 1) errors of incorrect sampling, 2) errors of correct sampling. Some incorrect sampling errors arise from what the GUM [2] refers to as gross errors, and as such would be excluded from estimates of uncertainty. The errors of correct sampling occur within good practice and can be considered for inclusion within estimates of uncertainty following the GUM approach [2].

10.2.4 *Incorrect sampling errors* arise from sampling equipment and procedures that do not follow the rules of sampling correctness defined in the sampling theory. In Figure 3 these errors are shown in shaded boxes. Increment delimitation error (IDE) is error that is generated if the shape of sample is not correct. For example, from a process stream the correct sample is a complete slice of equal thickness cut through the process stream. The sampling device should be designed so that it can extract the intended sample profile (i.e. all constituents have an equal chance to end up in the sample). Otherwise sample or increment extraction error

(IXE) is created. Sample preparation errors (IPE) have several potential causes listed in Table 2, two of which are excluded as gross errors by the GUM definition.

10.2.5 Incorrect sampling errors have the following properties in common: 1) they create sampling bias and increase the total variance in an unpredictable way, 2) they are circumstantial and, therefore, any attempt to estimate them experimentally is normally not useful, because it is expensive and the results cannot be generalised. The correct way is to minimise or eliminate these errors by carefully auditing the equipment and procedures, by replacing structurally incorrect devices and procedures with those that follow the rules of sampling correctness, and by sufficient training of sampling personnel. Only if this technical part is correctly executed does the theoretical part of uncertainty evaluation have predictive value. However, sampling uncertainty estimation and quality control may alert users to procedures that are not behaving correctly.

10.2.6 Correct sampling errors are shown in the lower part of Figure 3. When the incorrect sampling errors are eliminated these errors can be modelled and used for estimating the uncertainty of sampling. The fundamental sampling error is among the most important and will be considered further here; others are discussed in Appendix C2.

10.2.7 Fundamental sampling error (FSE) is the minimum error of an ideal sampling procedure. Ultimately it depends on the number of critical particles in the samples. For homogeneous gases and liquids it is very small but for solids, powders and particulate materials, especially at low concentrations of critical particles, fundamental error can be very large. If the lot to be sampled can be treated as a one-dimensional object, fundamental sampling error models can be used to estimate the uncertainty of the sampling. If the lot cannot be treated as a one-dimensional object, at least the point selection error has to be taken into account, when the variance of primary samples is estimated. If the sample preparation and size reduction by splitting are carried out correctly, fundamental sampling error models can be used for estimating the variance components generated by these steps. If the expectance value for the number of critical particles in the sample can be estimated easily as a function of sample size, Poisson distribution or binomial distribution can be used as sampling models to estimate the uncertainty of the sample. In most cases the fundamental sampling error model can be used.

10.2.8 If the material to be sampled consists of particles having different shapes and size distributions it is difficult to estimate the number of critical particles in the sample. An equation can be used to estimate the relative variance of the fundamental sampling error:

$$\sigma_r^2 = Cd^3\left(\frac{1}{M_S} - \frac{1}{M_L}\right) \qquad \text{Equation 5}$$

where

$\sigma_r = \frac{\sigma_a}{a_L}$ = relative standard deviation of the fundamental sampling error

σ_a = absolute standard deviation (in concentration units)

a_L = average concentration of the lot

d = characteristic particle size = 95% upper limit of the size distribution

M_S = sample size

M_L = lot size

C is a sampling constant that depends on the properties of the material sampled; C is the product of four parameters:

$$C = f\,g\,\beta\,c \qquad \text{Equation 6}$$

f = shape factor (see Figure 4)

g = size distribution factor (g = 0.25 for wide size distribution and g = 1 for uniform particle sizes)

β = liberation factor (see Figure 4). For materials where the particles are completely liberated, β = 1. For unliberated material an empirical equation, $\beta = \left(\frac{L}{d}\right)^x$, is used, where values of x ranging from 0.5 to 1.5 have been suggested.

c = constitution factor and can be estimated if the necessary material properties are available by using:

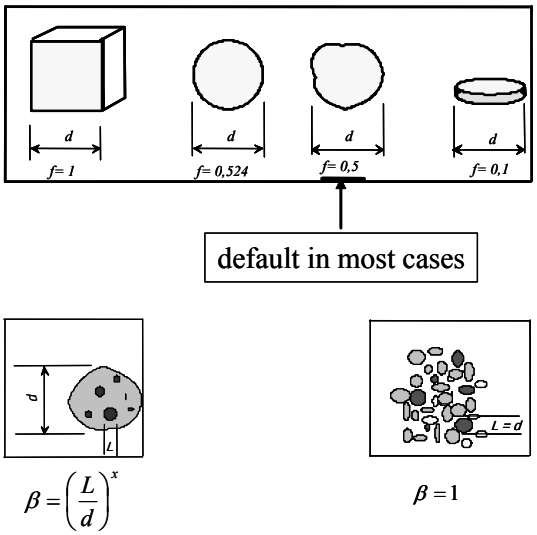
$$c = \frac{\left(1-\frac{a_L}{\alpha}\right)^2}{\frac{a_L}{\alpha}}\rho_c + \left(1-\frac{a_L}{\alpha}\right)\rho_m \qquad \text{Equation 7}$$

Here a_L is the average concentration of the lot, α the concentration of the analyte in the critical particles, ρ_c the density of the critical particles and ρ_m the density of the matrix or diluent particles.

10.2.9 If the material properties are not available and they are difficult to estimate, the sampling constant C can always be estimated experimentally. Certified reference materials, for example, are a special group of materials for which the sampling constant can be estimated from existing data.

10.2.10 An example of how the fundamental sampling error model can be used in practice is given in Example A5, Appendix A.

Figure 4 Estimation of factors for the estimation of fundamental sampling error.



The particle shape factor f (upper part), and liberation factor β for unliberated material (lower left) and liberated material (lower right). L is the liberation size of the critical particles.

PART 4 – Management issues

11 Responsibility for quality of sampling

11.1 The implications of regarding sampling as an integral part of the measurement process are far reaching, and include management issues. The rigour that is applied to assessing and improving the quality of activities within the analytical laboratory should be applied equally to the sampling procedures. The responsibility for the quality of the whole measurement process should ultimately rest with one organisation, and responsibilities for different parts of the process must additionally be defined. Similarly, one body should take responsibility for estimating the measurement uncertainty, based on information from all participants. This organisation can then inform all of the participants of the contributions arising from the main steps in the measurement procedure.

12 Selection of uncertainty estimation approach

12.1 The empirical (top-down) and modelling (bottom-up) approaches each have their advantages in certain circumstances. These should be considered in selecting the approach for a particular sampling exercise.

- The empirical approach includes all sources of uncertainty, without the scientist having to know their identity in advance. For example, it is automatically applicable to the particular contaminants and mineralogy present at a geological site. The calculations do not require any prior knowledge of the nature of the material (e.g. grain size, analyte speciation, degree of heterogeneity). It is relatively quick and easy to apply practically (especially for the 'duplicate method'). There are at least four options available to allow progressively more accurate (and more expensive) estimates of uncertainty, as appropriate (Table 5). Some of these methods can allow for systematic error (such as sampling bias) within the estimate of uncertainty. Sampling proficiency tests and reference sampling targets are still in the early stages of development, but already show considerable promise for this application.
- Among the disadvantages of the empirical approach is that it does not necessarily quantify any of the individual components of uncertainty (although this knowledge can be added with limited resolution). It is not based on a theoretical model of particulate sampling, but this may be an advantage in applications to materials that are not particulate in form (e.g. gaseous, liquids, biota). The empirical approach only gives an approximate value of uncertainty, which is assumed to be constant over the target, but this is also true of the modelling approach. Extreme values in the replicate measurements may lead to an overestimate of the uncertainty value, which is not representative of most measurements. This effect can be minimised, however, by the use of robust statistics.
- The principal advantage of the modelling approach is that it allows the largest source of uncertainty to be readily identified, if it was in the model. It gives a transparent method showing which components of uncertainty have been considered in the summation of uncertainty. Finally, where prior information is available, modelling approaches can be less costly than extensive experimental studies.
- The disadvantages of the modelling approach include that the theoretical predictions of uncertainty may require detailed prior measurements of the mineralogy, grain size and analyte speciation of the material to be sampled (e.g. soil), and how these vary across the

target. Idealised assumptions have to be made therefore about the make up of the material (e.g. mineralogy, grain size and analyte speciation). The modelling approach using sampling theory requires estimates or assumptions about eight types of sampling error, and also how these might vary across the target. Both theoretical and empirical approaches can be relatively time consuming and therefore expensive to implement. Generic estimates may be too general and not reflect the specific circumstances at any particular sampling target. Further, not all of the sources of uncertainty might be identified, leading to an underestimate of the total uncertainty.

On balance, therefore, the empirical methods tend to be more generally applicable across a wide range of types of material, and do not depend as heavily on prior knowledge of the system or all of the sources of uncertainty. This will make them less time consuming, and therefore less costly to apply, which is particularly valuable in one-off testing of different sampling targets. The modelling approaches, by contrast, lead to a more detailed assessment of individual known sources of uncertainty and are more appropriate when developing a long-term sampling scheme for a specific well-characterised application.

13 Quality control of sampling

13.1 Relationship between validation and quality control

13.1.1 Once an uncertainty that makes the measurements fit for purpose has been established, evaluation of the sampling and analytical procedures proposed to meet those purposes can be undertaken. Two evaluation tools are needed for this purpose: validation and continual quality control.

13.1.2 Validation comprises a one-time estimation of the uncertainty components determined under conditions expected to be encountered in the routine use of the procedures. The validation may be done generically for the sampling method (initial validation) or site specifically for the method used ‘on site’ to the selected target (on-site validation). Initial validation is used when sampling is done as a one-time campaign (spot sampling, e.g. contaminated site investigation) and on-site validation is repeated at intervals (repeated sampling, e.g. time or flow-proportional sampling of waste water). In short, validation demonstrates what can be achieved and, if that conforms to the fitness-for-purpose requirement, the procedures are deemed suitable for routine use. The methods for validation are described in the previous chapters of this Guide.

13.1.3 Validation alone cannot ensure that routine results are indeed fit for purpose, however. Routine or site specific conditions may differ from those prevailing during validation, either systematically or occasionally. This is especially true for sampling, where the larger part of the uncertainty component often stems from the heterogeneity of the target, that is, where the degree of heterogeneity may vary markedly from one target to the next. This is also true when a sampling method is applied at different sites. These circumstances emphasise the need for an ongoing internal quality control that includes sampling, to ensure that conditions prevailing at validation (and therefore the expected uncertainty attached to the results) are still applicable every time that the sampling and analytical procedures are executed. The combined use of validation and quality control is shown in Table 6.

Table 6: Illustration of the combined use of validation and quality control of sampling

	One method used at many sites	One method used repeatedly at one site
Validation	Initial validation yielding generic performance data	On-site validation yielding the performance data for the specific target
Quality control	Extensive quality control with site specific verification of generic performance data	Spot quality control verifying the performance data consistency over time

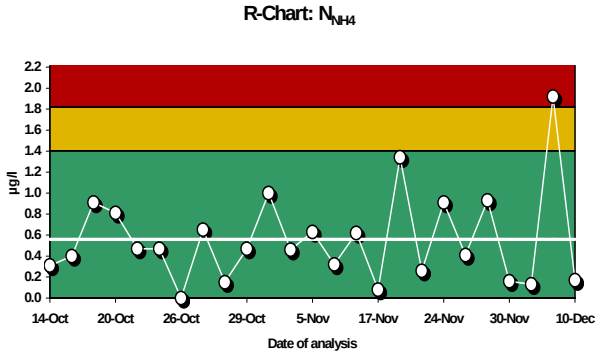
13.1.4 The need for internal quality control of sampling is not widely recognised at present, and methods for executing it are not well established, except in some specialised areas such as geochemical prospecting [21]. Specific suggestions for sampling quality control are given for some environmental sampling matrices in [22]. However, no new principles are involved; with minor qualification, the principles of internal quality control of analysis are applicable to sampling [23, 24, 25]. Moreover, the methods used in validation are, with some simplification, applicable to internal quality control. The reason for the simplification is that validation needs to provide a good estimate of uncertainty, while quality control merely needs to demonstrate consistency, over space and time, with the uncertainty established at validation.

13.2 Methods of internal quality control of sampling

13.2.1 The focus of interest is almost exclusively the precision aspect. Bias is difficult to address in validation and almost impossible in internal quality control. The 'reference target', the conceptual equivalent in sampling of a certified reference material [26], is rarely available. Moreover, it is not fully useful: we need to see whether results for individual sampling targets are fit for purpose, not whether unbiased and reproducible results can be obtained on a possibly unrepresentative reference target.

13.2.2 The principal tool is replication. This is minimally executed by taking two samples from each target by a complete (and suitably randomised) duplication of the sampling protocol. Each sample is analysed once and the difference between the results $D = \left| x_1 - x_2 \right|$ calculated. If the validated uncertainties of sampling and analysis are u_s and u_a respectively, the combined standard uncertainty is $u_{meas} = \sqrt{\left(u_s^2 + u_a^2 \right)}$. Consequently, a one-sided range control chart can be constructed with a control limit (at the 95% confidence interval) of $2.83 u_{meas}$ and an action limit (at the 99% confidence interval) of $3.69 u_{meas}$ [25] (Figure 5). An out-of-control value of *d* shows that the result should be scrutinised as possibly unfit for purpose. Such a result is not diagnostic and may stem from a disturbance in either sampling or analysis; the latter should be detected by standard methods of analytical quality control.

Figure 5: Example of an R-chart for quality control of sampling



For the construction of this R-chart see the Nordtest Guide [25].

13.2.3 The data from quality control can also be used to update sampling method precision as obtained in method validation using the same methods, ANOVA or relative difference calculations.

13.2.4 In some instances, the extra cost of duplicate sampling can be eliminated by use of the SAD (Split Absolute Difference) method in which the normal number of increments to be combined as the sample is segregated at random into two equal sized sub-sets, each of which is processed and analysed separately [27, 28]. The difference between such results has an uncertainty of $\sqrt{4u_s^2 + 2u_a^2}$ if conditions applying to validation are maintained. This again could be used to define an action limit in a one-sided control chart.

14 Reporting and interpreting uncertainty

14.1 Introduction

14.1.1 It is crucial to ensure that reports are clear as to the measurand being reported. In particular, it is important to be clear whether the result and its uncertainty apply to a single test portion, a laboratory sample, the whole of a sampling target (e.g. a bulk material), or a series of targets. Using the principles of the GUM [2] and previous Eurachem/CITAC Guides [1], uncertainty will initially be estimated in the form of a standard uncertainty, u, which includes due allowance for all effects which may reasonably influence the result. Uncertainty may be quoted in this form without change. However, it is often convenient to report in other forms for increased confidence or for wider applicability. It is essential to note any limitations in the estimate of uncertainty, such as the exclusion of sampling bias or other neglected effects. The following paragraphs describe the most important issues, and give some guidance on their interpretation.

14.2 Expanded uncertainty, *U*

14.2.1 The standard uncertainty u applied to a result in the form $x \pm u$, and associated with a normal distribution, describes an interval including only about 68% of the area of the distribution. This is usually taken to indicate that there is a greater than 32% probability of the measurand value being outside this interval. This is considered insufficient confidence for most practical applications. It is therefore normal practice to apply a suitable multiplier to the standard uncertainty so that the quoted interval includes a greater proportion of the dispersion.

Conventionally, this multiplier, usually designated k, is referred to as the *coverage factor*, and the product $ku=U$ is referred to as the *expanded uncertainty*.

14.2.2 The choice of k is discussed in considerable detail in other publications [1, 2]. However, the key principles are:

- k should be chosen to reflect an approximate confidence interval for the particular distribution.
- If a particular distribution is known to be applicable, it is used. Otherwise, a normal distribution is considered a reasonable assumption where the dominant contributions to uncertainty are all normally distributed or there are several approximately equal contributions from arbitrary distributions. With this assumption, k is typically based on the value of Student's t for an appropriate (two-tailed) level of confidence and number of degrees of freedom.
- In the modelling approach, the number of degrees of freedom is formally derived from the degrees of freedom for contributing uncertainties according to a published formula [1, 2], or approximated from the number of degrees of freedom for the dominant contribution [1]. More commonly, the number of degrees of freedom is assumed to be sufficiently large to justify a choice of k=2 for approximately 95% confidence.

For most practical purposes, k=2 is considered acceptable, and is sometimes mandatory [29]. However, it is important to state the value of k used, and the approximate level of confidence that k implies, when reporting expanded uncertainty.

14.3 Relative uncertainty statements

14.3.1 It is often found that the standard uncertainty from sampling increases approximately proportionally with the value of the result. Under these circumstances, it is often most practical to quote the uncertainty in a relative form, such as a relative standard deviation (u/x) or percentage interval using Equation 4 (e.g. ±10%). The relative value quoted is usually based on an estimate of uncertainty for one or more representative results, but is applicable over a greater range of concentration values.

14.3.2 It is important not to extrapolate a simple relative standard deviation to zero concentration, as uncertainty does not normally disappear entirely at very low levels and the proportionality assumption is no longer valid. More general approaches to these situations can either specify a range of concentration over which the relative uncertainty value applies [25], or else express the uncertainty as a function of concentration [1, 42].

14.4 Contributions to uncertainty

14.4.1 The exact steps that are included in each contribution to the measurement uncertainty need to be stated. Depending on the method of estimation employed, the details of the experimental design, and the person to whom the information is intended, it is possible to quantify some specific components of the measurement uncertainty. For example, the experimental design in Figure 2 will give separate estimates of two components called 'sampling' and 'analysis'. When the details of this particular implementation of the design are examined it becomes evident that uncertainty from physical sample preparation is included within the general title of 'sampling', whereas that from chemical preparation is included within 'analysis'. If required, it is possible to insert a further level of duplication of physical preparation within the experimental design to estimate the separate contribution which that particular step introduces [30]. The exact steps that are included in each contribution to the measurement uncertainty need to be documented. For less experienced users of analytical

measurements, it may be better to report one value for the whole uncertainty of the measurement, stating in a footnote which sources have been considered.

14.5 Applicability of estimates

14.5.1 Recalling the discussion of specification of the measurand (Section 5.2) it is crucial to ensure that reports are clear as to the measurand being reported. As observed in Section 14.1.1, it is particularly important to state clearly whether the result and its uncertainty apply to a single test portion, a laboratory sample, the whole sampling target, or to a series of targets. Unlike estimates of uncertainty for analytical measurements, it is very probable that the same sampling protocol will produce measurements with different levels of uncertainty from sampling when it is applied to a new sampling target. New estimates will be required for substantially different targets, particularly when there is reason to suppose that the degree of heterogeneity has changed.

14.6 Interpretation of uncertainty statements against limits

14.6.1 Results are often compared with tolerances or regulatory limits in order to assess compliance with a requirement. In making such comparisons, it is important to take uncertainty into account. A full discussion is beyond the scope of the present Guide; more detailed discussion will be found in references [1] and [31]. The basic principles are:

- Decide whether the decision requires proof of compliance, proof of *non*-compliance, or a 'shared risk' approach, and set an appropriate level of confidence.
- For proof of compliance, the result and its uncertainty interval must be entirely within the permitted range.
- For proof of *non*-compliance, the result and its uncertainty interval must be entirely *out*side the permitted range.
- For shared risk approaches, set a range for acceptable measurement results based on the permitted interval, adjusted to provide a specified probability of false acceptance and false rejection rates. Recent guidance gives useful details of the procedure [32].

For regulatory purposes, it is important to consult the specific regulations applicable, as no general guidance can currently cover all cases. For example, it is generally considered unsafe to 'pass' material that is not proven compliant, dictating a proof of compliance approach. Criminal prosecution in most countries, however, requires clear proof of *non*-compliance and in these circumstances (e.g. blood alcohol prosecutions) it is normal practice to seek proof of non-compliance at high levels of confidence.

15 Cost of estimating uncertainty from sampling

15.1 It would seem logical to consider the total budget for validation and quality control of sampling to be judged together against the costs that will arise from erroneous decisions based on inadequate estimates of uncertainty. It is recognised that implementing uncertainty estimation will increase the overall costs of measurement. Applying the duplicate method, for example, can increase the cost of sampling by up to 10%, and the analysis by 30% (i.e. three additional analyses are required for applying the balanced design to 10% of the sampling targets. An unbalanced experimental design may be performed, in which only one of the duplicate samples is analysed twice, if suitable statistical treatment is performed. This increased cost can be justified, however, by the additional information gained and the reduced

potential losses from incorrect decisions that might have been made without knowledge of the uncertainty (Section 16).

15.2 It is more difficult to evaluate general costs for the other methods of uncertainty estimation. Inter-organisational sampling trials require the expenses of at least eight different participants (to obtain an acceptable reliability [11]), and are therefore likely to be significantly higher than those for the duplicate method. Modelling methods will require detailed information about the material being sampled. For some materials that are relatively consistent over many batches these values may be generally applicable, and therefore make this approach more cost-effective than empirical methods that take larger numbers of extra measurements on each batch. This discussion must therefore include the extent to which the uncertainty value for a particular protocol/material combination is estimated at a preliminary validation, and how much the value is continually monitored and/or updated by an ongoing sampling quality control scheme (Section 1). It would seem logical to consider the total budget for validation and quality control of sampling to be judged together against the costs that will arise from erroneous decisions based on inadequate estimates of uncertainty.

16 Judging fitness for purpose of measurements using uncertainty

16.1 A proper understanding of uncertainty from sampling must be embedded in the broader perspective of fitness for purpose. Three approaches have been suggested for setting fitness-for-purpose criteria. The first approach is to set an arbitrary limit on the maximum value of uncertainty that is considered acceptable. This approach has been widely applied in the analytical sector, where a target relative uncertainty has been applied (e.g. 10%). The problem with this approach is that it does not necessarily relate to intended purpose for which the user requires the measurement.

16.2 The second approach is to compare the variance generated by the measurement (sampling and analysis) to the variance of the measurements between the different sampling targets. There are many situations where the objective of the measurements is to compare concentrations between different targets, such as in mineral exploration where the objective is to locate a target with significantly higher concentration of an element of interest (e.g. gold). One application of this approach, for example, sets the fitness-for-purpose criterion so that the measurement variance does not contribute more than 20% to the total variance (defined in Equation 2) [33].

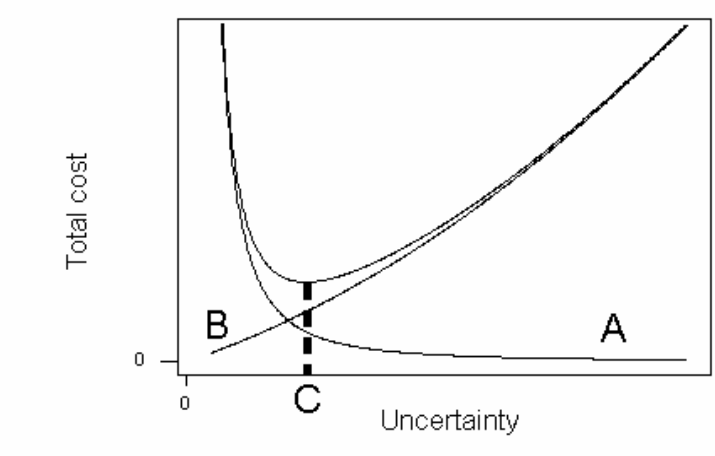
16.3 The third, and most generally applicable, approach to judging the fitness for purpose of measurements, is consider the effect of the measurement on its ultimate purpose. All analytical measurement is undertaken to support a decision. A decision can be either correct or incorrect. An incorrect decision involves extra costs, and an incorrect decision is more likely if the uncertainty is higher. Consider, for example, the manufacture of a material against a specification of a maximum acceptable level of an impurity.[g] Each batch of material is analysed to determine the level of the impurity. A ‘false positive’ result has the outcome that the batch of material is discarded or reworked unnecessarily to reduce the apparently unacceptable level of impurity. A ‘false negative’ result means that a defective batch is released to the customer, a situation that may require financial compensation. Both of these situations are more likely to occur if the uncertainty is higher. This seems to suggest that the measurement should be undertaken so that the uncertainty is the smallest that can be

[g] This concept is equally applicable to situations where materials have regulated minimum analyte concentrations, in which case the terms ‘false compliance’ and ‘false non-compliance’ are applicable.

achieved. However, reducing the uncertainty of a measurement result requires rapidly escalating costs. A useful rule here is that, where random variation dominates the uncertainty, the cost of a measurement is inversely proportional to the square of the uncertainty; a reduction in uncertainty by a factor of 2 calls for an increase in expenditure by a factor of 4.

16.4 The true cost of a decision is the sum of the measurement costs and the excess costs of incorrect decisions. From the above we can see that this sum has a minimum value at some particular level of uncertainty (Figure 6), and this uncertainty is the definition of fitness for purpose.

Figure 6: Schematic loss functions dependent on uncertainty of measurement



Line A shows the costs of measurement. Line B shows costs of incorrect decisions. The sum of these two lines (the total cost shown by the highest line) shows a minimum cost at point C, which is the uncertainty that can be regarded as fit for purpose.

16.5 The optimal apportionment of resources between sampling and analysis is also a matter of costs. Even an elementary consideration (excluding costs) shows that the uncertainties of sampling and analysis should be roughly balanced. For example, if the uncertainties of sampling and analysis are 10 and 3 units respectively, the overall uncertainty of measurement is $\sqrt{10^2 + 3^2} = 10.4$. The overall uncertainty is hardly affected by a reduction of the uncertainty of analysis: if it is reduced to (say) 1 unit, the overall uncertainty is reduced to $\sqrt{10^2 + 1^2} = 10.05$, an inconsequential change. A more sophisticated approach takes into account the different costs of analysis and sampling. If the unit costs of sampling and analysis are A and B for the same specific level of uncertainty, the optimum ratio of sampling uncertainty u_{samp} to analytical uncertainty u_{anal} is given by

$$\frac{u_{samp}}{u_{anal}} = \left(\frac{A}{B}\right)^{1/4}.$$

This ratio provides the minimum expenditure for a given overall uncertainty of $\sqrt{u_{samp}^2 + u_{anal}^2}$ or, alternatively, the minimum uncertainty for a given expenditure [34].

Methods for modifying uncertainty from sampling are discussed in Appendix E, although operating at 'minimum total cost' is not always achievable or necessary.

17 Implications for planning sampling and measurement strategies

17.1 Expertise and consultation

As Section 4 shows, the sampling and analytical processes cover a range of activities. Different parts of the process are frequently allocated to different staff, who may have very different knowledge of the objectives and, more importantly, differing knowledge of the effect of different parts of the process. In general, all of those involved will have good knowledge of some part of the process, but few are able to advise on the complete process. It is therefore important that sample planners involve analytical chemists and experienced sampling technicians where possible in planning sampling. It is also prudent to include statistical experts in most circumstances (see below). Decision makers (i.e. business managers and those acting on the results of sampling activities) should be involved in planning for new applications, and regulators should also be consulted where a protocol is intended to support regulation.

Although the principles of this Guide are widely applicable, expert statistical guidance is always valuable and should be considered essential in some circumstances. These include:

- where the observed or expected frequency distributions are not normal, for example where the results contain more than 10% outliers, or where the results show markedly asymmetric distributions;
- where large financial or social consequences depend on a reliable estimate of uncertainty;
- where confidence intervals are needed on the estimates of uncertainty or, for more complex sampling plans, on the measurement results;
- where the sampling strategy is more complex than simple random sampling with replicated measurements, for example in implementing stratified sampling.

17.2 Avoiding sampling bias

The methods described in this Guide are suitable for establishing the variability of sampling, but only the more complex methods can begin to assess uncertainties associated with possible bias in sampling. For this reason, close attention should be paid to minimising potential sources of bias. These include possible bias associated with differential sampling due to particle size, density or flow-rate; bias in selection of sampling points; the effect of different sampling equipment etc. Specific expertise in sampling methodology should be sought unless these factors can be demonstrated to be adequately controlled or are completely specified by an established sampling protocol.

17.3 Planning for uncertainty estimation

Sampling exercises should always make provision for at least some replicated samples and measurements in order to assess the uncertainty of the results.

17.4 Fitness-for-purpose criteria

Planning should ideally begin with the establishment of clear fitness-for-purpose criteria, taking into account the relative costs and uncertainties of sampling and analysis where they are known or can reasonably be determined in advance. Section 16 provides guidance on how analytical and sampling effort can be optimised.

17.5 Use of prior validation data

The main uncertainties associated with analytical measurements are often estimated during, or on the basis of, analytical method validation, a process which is carried out prior to bringing the method into use. Consideration accordingly needs to be given as to whether the variability found as part of the sampling experiment should replace, inform, or simply serve as a check on, the analytical measurement uncertainty assessed using prior information. In considering this issue, it should be noted that the variability observed during a relatively short series of analyses is rarely sufficient as an estimate of uncertainty. Long-term studies are generally more reliable. It is accordingly safer to rely on prior validation data unless the observed variation is significantly higher.

Uncertainties associated with sampling variability can themselves be estimated in advance, particularly where a long-term sampling programme is to be planned and implemented. Under these circumstances, it is usually prudent to obtain an initial estimate of sampling uncertainty. Ongoing studies can then serve as a check on continuing validity of the uncertainty estimate, for example by applying internal quality control principles as discussed in Section 13.

17.6 Acceptability of sampling uncertainty

Before reporting measurements, it should be evaluated whether they are acceptable and in accordance with the quality objectives set for the whole uncertainty and its sampling component, probably based on some fitness-for-purpose criterion, prior to the measurements.

Appendix A: Examples

Introduction

The most effective way to explain the methodologies described in the main text of this Guide is to show worked examples. These examples are not intended to cover all circumstances, but to show how the general principles can be applied to a variety of situations across a range of different sectors. These include food (production and retail), animal feed, and environment (soil and water). The examples are all structured using the same basic format, so as to aid comprehension and comparability.

Example A1: Nitrate in glasshouse grown lettuce

Measurand				Uncertainty estimation		
Analyte/ Technique	Unit[1]	Sector/ Matrix	Sampling target(s)	Purpose	Design	Statistics
Nitrate/Hot water extraction and determination by HPLC	mg kg^{-1} as received	Food/ Lettuce	1 bay of Iceberg lettuce grown under glass	Uncertainty – total measurement, sampling and analytical	Empirical - duplicate method	Robust ANOVA

1 Scope

Estimate the measurement uncertainty, and contributions from sampling and analysis, for routine monitoring of glasshouse grown lettuce, using a standard sampling protocol.

2 Scenario and sampling target

Nitrate is essential for plant health; however, there are concerns for human health associated with eating elevated levels of nitrate. The concentrations of nitrate in lettuce are regularly monitored in line with EC requirements. Concentration estimates are made for each 'bay' of up to 20,000 lettuce heads, and the result for each bay used individually in assessing conformance with the relevant Regulation. Each bay is accordingly considered a sampling target, rather than individual heads of lettuce. In order to make a reliable comparison of the measured nitrate concentrations against the European regulatory threshold [35] (4500 mg kg^{-1}), an estimate of the measurement uncertainty is desirable.

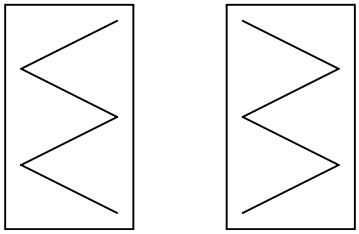
3 Sampling protocol

The accepted protocol for this purpose specifies that one composite sample is prepared from 10 heads of lettuce harvested from each bay of lettuce [36]. The lettuces are selected by walking a W shape or five-point die shape through the bay under investigation. This protocol is applied to all bays regardless of the size. Samples were taken in the morning and transported to the contracted analytical laboratory in ice-packed cool boxes to arrive within 24 hours of sampling.

4 Study design – duplicate method (Section 9.4.2)

The minimum of eight targets were selected for inclusion in the uncertainty estimation protocol. For each of these bays a second 10-head sample was taken (S2) in addition to the routine sample (S1). This duplicate sample was taken in a way that represented the variation that could occur due to the ambiguities in sampling protocol, for example positioning of the origin of the W design, and its orientation.

Figure A1.1: Example of the 'duplicate method'



Example of how the duplicate method can be applied. Using the W design as an example, the protocol stipulates the design but not the position or orientation. The 'W' is equally likely to start on the left or the right. Ten heads are taken along the line of the W to create a composite sample for one target.

5 Sampling and analysis in the laboratory

Primary samples were frozen on receipt at the laboratory. A lettuce (increment) from each 10-head sample was cut into four equal quarters and two quarters retained. This was repeated for each of the 10 increments in the sample. The resultant 20 quarters were place in a Hobart processor and macerated to process a composite sample. Two analytical test portions (10 g) were taken. Each test portion was extracted using hot water and the nitrate concentration was determined by HPLC (ultra-violet detector). Quality control samples (spike recovery) were analysed concurrently with the real samples. No significant analytical bias could be detected and so bias correction was considered unnecessary for the resultant data. The raw measurement values use for the estimation of uncertainty had appropriate rounding, and no suppression of values less than either zero, or the detection limit.

6 Results

The best estimates of the nitrate concentration at each of the eight target locations are shown in Table A1.1.

Table A1.1: Measurements of the concentration (mg kg^{-1}) of nitrate in eight duplicated samples. The duplicate samples are labelled S1 and S2. Likewise, duplicate analyses are labelled A1 and A2. Hence, DS1A2 (value 4754 mg kg^{-1}) is analysis 2, from sample 1 from sampling target D

Sample target	S1A1	S1A2	S2A1	S2A2
A	3898	4139	4466	4693
B	3910	3993	4201	4126
C	5708	5903	4061	3782
D	5028	4754	5450	5416
E	4640	4401	4248	4191
F	5182	5023	4662	4839
G	3028	3224	3023	2901
H	3966	4283	4131	3788

Before applying statistical methods it is useful to inspect the data to ascertain the general levels of variability. The analytical duplicates (e.g. BS1A1 and BS1A2) are generally within 300 mg kg^{-1} of each other, suggesting an analytical precision of less than 10%. The sample duplicates (e.g. DS1 and DS2) agree less well, but generally differ by less than 20%. However, one target (C) displays a greater difference, suggesting an outlying value.

Quantification of the random component of the measurement uncertainty and two of its main components (sampling and analysis) was made using robust analysis of variance (RANOVA, Appendix C3, with output in Figure.A1.2. Robust ANOVA was used here as outlying targets are, in this relatively well-controlled environment, considered likely to be anomalies, rather than reflecting the underlying population statistics, and as a precaution against analytical outliers.

Note: Robust methods should *not* be used where apparent outliers arise as part of the typical population of sampled increments or targets, unless the specific implementation allows for non-normal distributions for part of the assumed error structure.

Figure A1.2

CLASSICAL ANOVA RESULTS

Mean = 4345.5625

Standard Deviation (Total) = 774.5296

Sums of Squares = 12577113 4471511 351320

	Between-target	Sampling	Analysis
Standard Deviation	556.2804	518.16089	148.18063
Percentage Variance	51.583582	44.756204	3.6602174

ROBUST ANOVA RESULTS:

Mean = 4408.3237

Standard Deviation (Total) = 670.57617

	Between-target	Sampling	Analysis	Measurement
Standard Deviation	565.39868	319.04834	167.94308	360.5506
Percentage Variance	71.090791	22.636889	6.2723172	28.909209
Relative Uncertainty (% at 95% confidence)	–	14.474814	7.6193626	16.357719

The output of ANOVA for data produced form a balanced experimental design (n = 8, Table A1.1). Both robust and classical estimates are given for comparison. Standard deviation estimates are computed for 'between-target' ($s_{between\text{-}target}$), 'within-target' ($s_{samp}$) and within-chemical analysis (s_{anal}). Results are in the same units of concentration as the input data (i.e. mg kg^{-1} in this case).

Extracting the robust estimates from this output gives:

s_{samp} = 319.05 mg kg^{-1}

s_{anal}= 167.94 mg kg^{-1}

Equation 1 can be used to calculate:

$s_{meas} = \sqrt{(s_{samp}^2 + s_{anal}^2)} = 360.55$ mg kg^{-1}

This can be used as an estimate of the random component of the standard uncertainty (u).

The expanded relative uncertainty is given by Equation 3 as:

U_{meas}' = 200 * 360.55 / 4408 = 16.4% (*of concentration value)*

For the sampling alone, the expanded relative uncertainty (random component) is similarly given by:

U_{samp}' = 200 * 319.05 / 4408 = 14.5%

For comparison the expanded uncertainty for the analytical contribution (random component) is given by:

U_{anal}' = 200 * 167.94 / 4408 = 7.6%

This value is less than the normal limits set within internal analytical quality control (e.g. 10%).

The analytical recovery estimates were not statistically different from 100% recovery (i.e. no analytical bias was detected). For this example, therefore, no additional allowance was made for uncertainty associated with analytical bias.

7 Comments

This uncertainty estimate does not include any estimate of the possible sampling bias.

8 Assessment of the fitness for purpose of these measurements

The fitness-for-purpose criterion used initially is that based on the percentage of total variance (Section 16.2). When using RANOVA the program computes how much the between-target, within-target (or sampling) and analytical variance contributes (as a percentage) to the total variance (Figure A1.2). For this study of nitrate in lettuce the maximum contribution to the total variance is from between-target variability (71.1%). By combining the sampling (22.6%) and analytical contributions (6.3%) it is clear that the combined measurement process contributes 28.9% of the total variance. This is marginally greater than the ideal of 20%. Of this measurement variance, sampling is the dominant factor, responsible for 78.2% of the measurement variance.

Fitness for purpose may also be assessed using the optimised uncertainty (OU) methodology. This method addresses fitness-for-purpose assessment with financial considerations (Section 16.3)[37]. In this case it can be shown that an increase from a 10-head to a 40-head composite sample is required to achieve fitness for purpose (Appendix E, and [38]).

9 Reporting and interpretation

For each bay of lettuce (sampling target), the nitrate concentration of the 10-head composite sample is compared to the threshold value (4500 mg kg^{-1}). Each nitrate concentration should be reported with the measurement uncertainty (16.4% of the measured value) Table A1.2. The interpretation of whether each batch exceeds a threshold value, based upon its measurement and associated uncertainty, depends on the wording of the appropriate regulation [32]

10 Summary

Measurement uncertainty		
Sampling	Analytical	Total
14.5%	7.6%	16.4%

Table A1.2

Sample target	S1A1	Expanded Uncertainty
A	3898	639.3
B	3910	641.2
C	5708	936.1
D	5028	824.6
E	4640	761.0
F	5182	849.8
G	3028	496.6
H	3966	650.4

The nitrate concentrations associated with S1A1 (routine sample) are shown with the associated measurement uncertainty (calculated from $U = 16.4\%$). As an example, Target F has a value of the measurand (or true value) between 4332 mg kg^{-1} and 6032 mg kg^{-1}.

Example A2: Lead in contaminated top soil

Measurand				Uncertainty estimation		
Analyte/ Technique	Unit	Sector/Matrix	Sampling target(s)	Purpose	Design	Statistics
Total lead /ICP-AES	mg kg^{-1} dry basis	Environmental/ Top soil	100 targets – each of area 30 m x 30 m, with depth of 0–150 mm	Uncertainty – total measurement, sampling and analytical	Empirical - duplicate method	Robust ANOVA

1 Scope

Estimate the measurement uncertainty, and contributions from sampling and analysis, at each of 100 different sampling targets within one site, using a common sampling protocol.

2 Scenario and sampling target

An investigation was made of a 9-hectare site, as part of the assessment of the land for potential housing development. The most important analyte element for human health risk assessment was found to be lead. In order to compare the concentration of lead in the soil with the national regulatory threshold limit (450 mg kg^{-1}), an estimate of the lead concentration and the measurement uncertainty was required for each of 100 sampling targets.

3 Sampling protocol

One hundred samples of top soil (nominal depth 0–150 mm) were taken with a hand auger (diameter 25 mm) at 100 locations. These locations were distributed on a regular grid with sample spacing of 30 m (Table A2.1), and therefore each is intended to represent an area 30 m by 30 m. The surveying was conducted with a measuring tape and compass.

4 Study design – duplicate method (Section 9.4.2)

Ten of the samples (i.e. 10% of the total number), at randomly selected locations, were sampled in duplicate using the balanced design (Figure 2). The duplicate samples were taken at a distance of 3 m from the original sample, in a random direction. This aims to reflect the ambiguity in the sampling protocol, the uncertainty in locating the sampling target (e.g. the surveying error) and also the effect of small-scale heterogeneity on the measured concentration within the specified target. Six soil certified reference materials (CRMs) were selected for analysis to estimate analytical bias over a range of concentration.

5 Sampling and analysis in the laboratory

Primary samples were oven dried overnight at 60 C, disaggregated, sieved to remove particles with a natural grain size greater than 2 mm (based upon the definition of soil). The sieved samples (<2 mm) were all ground (95% < 100μm) and mixed. Test portions of 0.25 g were

taken for dissolution with nitric and perchloric acids, prior to determination of lead by ICP-AES [39]. The measurements were subject to full analytical quality control (AQC), and corrected for reagent blank concentrations where these values were statistically different from zero. The raw measurement values use for the estimation of uncertainty had no rounding or suppression of values less than either zero, or the detection limit.

6 Results

The best estimates of the lead concentration at each of the 100 target locations are shown in the format of a map (Table A2.1).

Table A2.1: Measured lead concentrations at each target on the sampling grid (mg kg^{-1}), shown by the actual coordinates used in the regular sampling grid (spacing 30 m). They show a high degree of variability between-locations of roughly a factor of 10. The variability within 10 of these locations selected at random (i.e. A4, B7, C1, D9, E8, F7, G7, H5, I9 and J5) was used for the estimation of uncertainty from sampling (Table A2.2). This within-target variation is substantial (e.g. a factor of 2) but substantially less than the between-target variability.

Row	A	B	C	D	E	F	G	H	I	J
1	474	287	250	338	212	458	713	125	77	168
2	378	3590	260	152	197	711	165	69	206	126
3	327	197	240	159	327	264	105	137	131	102
4	787	207	197	87	254	1840	78	102	71	107
5	395	165	188	344	314	302	284	89	87	83
6	453	371	155	462	258	245	237	173	152	83
7	72	470	194	82.5	162	441	199	326	290	164
8	71	101	108	521	218	327	540	132	258	246
9	72	188	104	463	482	228	135	285	181	146
10	89	366	495	779	60	206	56	135	137	149

Four measurements from the balanced design for each of the 10 sample targets selected for duplication were used for the estimation of uncertainty (Table A2.2). Visual inspection of the data allows an initial qualitative assessment of the relative importance of the two sources of measurement uncertainty. The low level of agreement between the concentration values from some of the sample duplicates is indicative of a high level of sampling uncertainty (e.g. S1 compared to S2 for target ‘D9’). The agreement between the analytical duplicates (A1 and A2) is however generally much better for most samples (< 10% difference) than that between the sample duplicates.

Table A2.2: Measurements of the concentration (mg kg^{-1}) of a lead on 10 duplicated samples from the total of 100 targets in a survey of contaminated land (Table A2.1). The duplicate samples are labelled S1 and S2. Likewise, duplicate analyses are labelled A1 and A2. Hence, D9S1A2 (value 702 mg kg^{-1}) is analysis 2, from sample 1 from sampling target D9. Values shown are rounded for clarity, and used for subsequent calculations, but generally un-rounded values are preferable for these calculations.

Sample target	S1A1	S1A2	S2A1	S2A2
A4	787	769	811	780
B7	338	327	651	563
C1	289	297	211	204
D9	662	702	238	246
E8	229	215	208	218
F7	346	374	525	520
G7	324	321	77	73
H5	56	61	116	120
I9	189	189	176	168
J5	61	61	91	119

Quantification of the random component of the measurement uncertainty and two of its main components (sampling and analysis) was made using robust analysis of variance (RANOVA), with typical output shown in Figure A2.1. Robust statistics were selected to allow for the outlying values that are evident in this data (e.g. target A4, sample duplicate D9S1/S2, analytical duplicate B7S2A1/A2), and in most similar data sets (but see the Note in Example A1, section 6). The estimates of uncertainty are averaged over the 10 targets, assuming that the uncertainty is not varying significantly over this range of concentration. The uncertainty is expressed in relative terms so that it is applicable over this range of concentration (Section 14.3).

Extracting the robust estimates from this output gives:

s_{samp} = 123.8 mg kg^{-1}

s_{anal} = 11.1 mg kg^{-1}

Equation 1 can be used to calculate:

$s_{meas} = \sqrt{(s_{samp}^2 + s_{anal}^2)}$ = 124.3 mg kg^{-1}

This can be used as an estimate of the random component of the standard uncertainty (u).

The expanded relative uncertainty is given by Equation 3, with a coverage factor of 2 as:

U_{meas}' = 200 * 124.3 / 297.3 = 83.63% (*of concentration value*)

For the sampling alone, the expanded relative uncertainty (random component) is similarly given by:

U_{samp}' = 200 * 123.8 / 297.3 = 83.29%

Figure A2.1: The output of ANOVA for data produced from a balanced experimental design (*n* = 10, Table A2.2)

CLASSICAL ANOVA RESULTS

Mean = 317.79999

Standard Deviation (Total) = 240.19238

Sums of Squares = 1738031.9 370075.5 6473

	Between-target	Sampling	Analysis
Standard Deviation	197.55196	135.43246	17.990274
Percentage Variance	67.646327	31.792678	0.5609926

ROBUST ANOVA RESULTS:

Mean = 297.30884

Standard Deviation (Total) = 218.48763

	Between-target	Sampling	Analysis	Measurement
Standard Deviation	179.67409	123.81386	11.144044	124.31436
Percentage Variance	67.62655	32.113293	0.26015487	32.373447
Relative Uncertainty (% at 95% confidence)	–	83.289726	7.4966113	83.626415

Both robust and classical estimates are given for comparison. Standard deviation estimates are computed for 'between-target' ($s_{between\text{-}target}$), 'within-target' ($s_{samp}$) and within-chemical analysis (s_{anal}). Results are in the same units of concentration as the input data (i.e. mg kg^{-1} in this case).

For comparison the expanded uncertainty for the analytical contribution (random component) is given by

$U_{anal}' = 200 * 11.1 / 297.3 = 7.5\%$

This value is less than the typical limits set within internal analytical quality control (e.g. 10%).

Inclusion of analytical bias

The analytical bias was estimated as –3.41% (±1.34%) using a linear functional relationship [40] established between the measured values on the certified values of the six CRMs (Table A2.3).

There is currently no consensus on the best way to combine random and systematic effects into an estimate of uncertainty, although four options have been identified [30]. One option [25] is to consider the estimated analytical bias (e.g. –3.41%) to be a typical value for participants in an inter-organisational trial. If this bias, and its own uncertainty (1.34%) is then added to the random component of the uncertainty (using the sum of squares) it will increase the variance to that which would be found in such a trial. The logic of this approach is that the extra uncertainty that is usually detected in inter-organisational trials is due to the unsuspected bias within each organisation. Where an estimate can be made of the extra variance caused by these biases between different laboratories, this can be added to the

random component within one organisation. In this case, the standard relative analytical uncertainty is increased to 5.24% [= $(3.75^2 + 3.41^2 + 1.34^2)^{0.5}$]. The expanded analytical uncertainty (10.48%) is then greater than the analytical target value of 10%, but it can also usefully be compared with an independent estimate of the analytical measurement uncertainty made within the laboratory. The expanded uncertainty for the whole measurement is thereby increased to 83.95% [= $(83.29^2 + 10.48^2)^{0.5}$], which is practically identical to the purely random component of 83.63%.

Table A2.3: Measured and certified lead concentration values for CRMs for the estimation of the bias of the analytical method

CRM name (n=4)	**Mean** (mg kg^{-1})	**Standard Deviation** (mg kg^{-1})	**Certified value** (mg kg^{-1})	***U* on certified value** **(95% conf.)**
NIST2709	19.7	3.2	18.9	0.5
NIST2710	5352.0	138.0	5532.0	80.0
NIST2711	1121.4	14.7	1162.0	31.0
BCR141	34.4	3.9	29.4	2.6
BCR142	36.2	4.6	37.8	1.9
BCR143	1297.5	33.0	1333.0	39.0

7 Comments

This estimate of uncertainty does not make allowance for any undetected sampling bias (Section 9.4.2). However, because the uncertainty is often dominated by the heterogeneity of the sampling target, the extra uncertainty introduced by bias in sampling can often be assumed to be insignificant by comparison (as shown for the analytical bias). Where the highest quality of uncertainty estimate is required, due perhaps to potentially large financial consequences from underestimating the uncertainty, it may be preferable to use one of the more elaborate methods using multiple samplers and/or protocols (Table 5).

If the measurand (or true value) had been defined as the mean concentration of lead across the whole site, the uncertainty would have had to include the contribution from the standard error on the calculated mean value, expressed as $s_{total} / \sqrt{n}$. For this example s_{total} is 403 mg kg^{-1}, n = 100 and the uncertainty on the mean (291.9 mg kg^{-1}) is therefore 27.6% of the value, at 95% confidence. This value can be calculated without knowing the individual contribution of the uncertainty from either the sampling or the analysis, and is often dominated by $s_{between\text{-}sample}$.

8 Assessment of the fitness for purpose of these measurements

Using the 'percentage of total variance' method (Section 16.2), the output in Figure A2.1 attributes the percentage of the total variance ([standard deviation (total)]^2 that is contributed by 'between-target', sampling (within-target) and analysis (within-sample). In this particular

example there is clearly a dominance of the 'between-target' variance (67.6% of total variance), although this is less than the ideal threshold of 80% (Section 16.2). Furthermore, sampling dominates (32.11% of total variance) over chemical analysis (0.26% of total variance) as a contributor to the measurement variance. Sampling variance (i.e. within-target) is identified as the principal contributor (99.2%) of uncertainty in the measurement process in this case (i.e. 100 * 32.11 / [32.11 + 0.26]).

The assessment of fitness for purpose of measurements in contaminated land investigation using the optimised uncertainty method (Section 16.3), is described elsewhere [41].

9 Reporting and interpretation

Individual measurements of lead concentration reported for these targets should have attached uncertainty values equal to 83.9% of the concentration value. This applies to all of these measured values (Table A2.1), which are at least 10 times higher than the analytical detection limit (estimated as 2 mg kg^{-1} in this case). In applications where this is not the case, it will be necessary to express the uncertainty as a function of the concentration [42]. Furthermore, the uncertainty on the mean measurements taken at the 10 targets where duplicate samples were taken (e.g. those listed in Table A2.2) will have reduced uncertainty estimates of 59.3% (= 83.9/√2).

Knowing the value of the uncertainty, it is also possible to make a probabilistic interpretation of the level of lead contamination on the site [43].

10 Summary

Measurement uncertainty*		
Sampling	Analytical	Total
83.3%	10.5%	83.9%

* with coverage factor of 2 (i.e. 95% confidence)

Example A3: Dissolved iron in groundwater

Measurand				Uncertainty estimation		
Analyte/ Technique	Unit	Sector/Matrix	Sampling target	Purpose	Design	Statistics
Dissolved iron/ ICP-AES	mg l^{-1}	Environment/ groundwater	The groundwater near one selected monitoring well in a groundwater body	Total uncer-tainty	Empirical duplicates used in validation and quality control	Range

1 Scope

The scope is determination of the total uncertainty of the measurement of dissolved iron in a sampling validation study and subsequent control of sampling uncertainty during monitoring.

2 Scenario and sampling target

A groundwater body that is an important drinking water resource for the city of Århus, the second largest city of Denmark, has through surveillance monitoring been identified as at risk for deterioration of the quality due to intensive drinking water abstraction. An operational monitoring programme has been established in order to control the trend in water quality development.

The groundwater body is in glacial outwash sand with Miocene sands and clays below and glacial till above. The geology at the site is complicated with several local aquifers (underground layer of water-bearing permeable rock, or permeable mixtures of unconsolidated materials) and aquitards (geological formation of layers comprised either of clay or on non-porous rock that restrict water flow from one aquifer to another). The groundwater body as identified is 2 km x 2 km x 10 m, starting 20–30 m below the surface. The natural quality of the groundwater is anaerobic without nitrate, with sulphate and reduced iron, but without hydrogen sulphide and methane. One of the threats to the groundwater body is oxygen intrusion into the aquifer as the result of the water abstraction and concomitant groundwater table drawdown.

In the groundwater body, nine wells had been sampled for chemical analysis during surveillance monitoring, and six wells are now available for sampling. In the operational monitoring plan, it was decided to aim at monitoring one well twice per year. The objective of the operational monitoring was set to having a 95% probability of recognising a 20% quality deterioration. It was decided to use dissolved iron as a target parameter that would be a sensitive indicator of aquifer oxidation (decreasing iron concentration with increasing oxidation) and with redox potential as supporting evidence. Oxygen, pH, electrical conductivity and redox potential were used as on-line indicators of sampling stability and sodium, calcium and chloride as general groundwater quality parameters. Only the two key parameters, dissolved iron and redox potential, are discussed here.

Meeting the monitoring objective requires a measurement uncertainty including both sampling and analysis of not more than 10% (comparison of two means each for two samples, 95% confidence interval, two-sided test) corresponding to an expanded measurement uncertainty of 20%. To ensure the compliance of the monitoring programme with the stated objective, a sampling validation study was initially conducted including all wells available and, based upon the results from this, a routine sampling quality control programme was set up for implementation with the monitoring programme for the selected monitoring well.

The properties of the groundwater body were summarised based upon previous monitoring activities (surveillance monitoring). Table A3.1 shows a summary for the two key parameters including variability in time and space as well as measurement (sampling and analytical) uncertainty.

Table A3.1: Key chemical parameters for nine wells to the groundwater body, surveillance monitoring

	Redox potential mV	**Dissolved iron** mg l^{-1}
Mean	-123	1.11
Relative standard deviation	27%	56%
Main cause of uncertainty	Oxygen impact during sampling and on-line measurement	Filtering

The chemical data suggest that the groundwater composition is quite uniform over time and space with respect to the main components (data not shown, relative standard deviation 1.9–16%), whereas the variability is high for the redox parameters (oxygen, redox potential and dissolved iron). The expected main causes of uncertainty are indicated in the table for the two key parameters and the causes were controlled during sampling.

3 Sampling protocol

Sampling was done according to the Århus County groundwater monitoring protocol, with permanent, dedicated pumps (Grundfos MP1) set in the middle of the screened interval of each well. Pump rates were 1–2 m^3 h^{-1} (well purging) with a 10% reduction just before sampling. Two of the six wells were large-diameter abstraction wells equipped with high yield pumps. These were pumped with 40–60 m^3 h^{-1} for well purging followed by pump rate reduction just before sampling. During well purging, the development in water quality was followed with on-line measurements of oxygen, pH, electrical conductivity and redox potential until stable readings and then samples were taken. A field report was filled in during the sampling, including pump yields and pumping times as well as water table measurements.

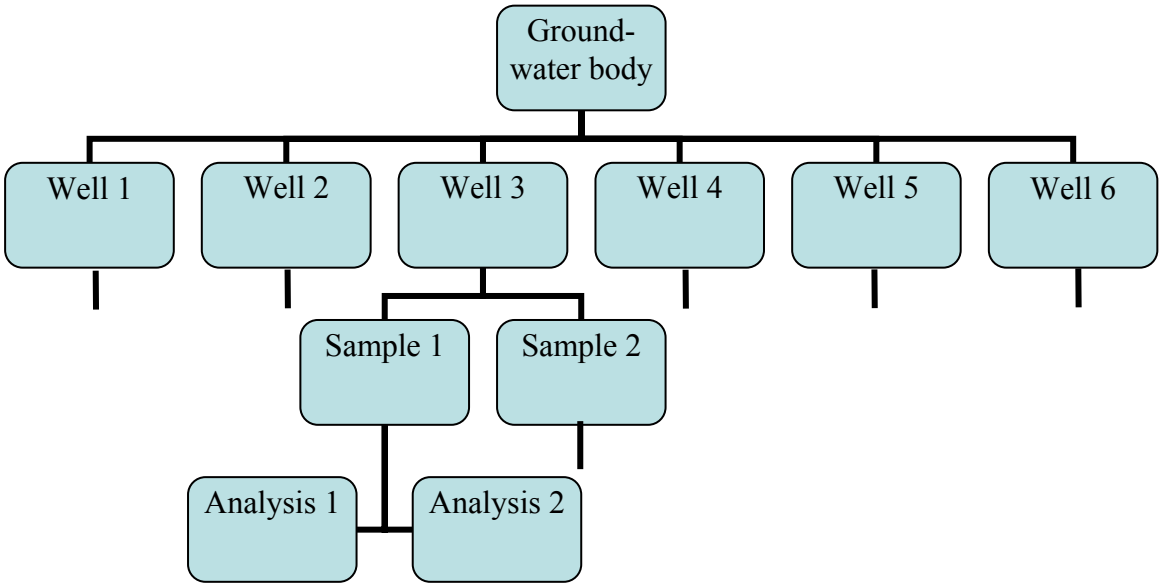
4 Study design – empirical

The empirical approach was selected for study design in order to provide estimates of heterogeneity in the groundwater body (between-target variation well-to-well and over time) and measurement uncertainty, split to show sampling uncertainty and analytical uncertainty.

4.1 Validation

The objective of the validation programme was to ensure that a measurement uncertainty meeting the set quality objective could be obtained and to describe the components of uncertainty in order to identify points of improvement, if required. The validation programme was set up with sampling of six wells, two independent samplings per well and two sub-samples per sample analysed, see Figure A3.1.

Figure A3.1 Design outline for validation

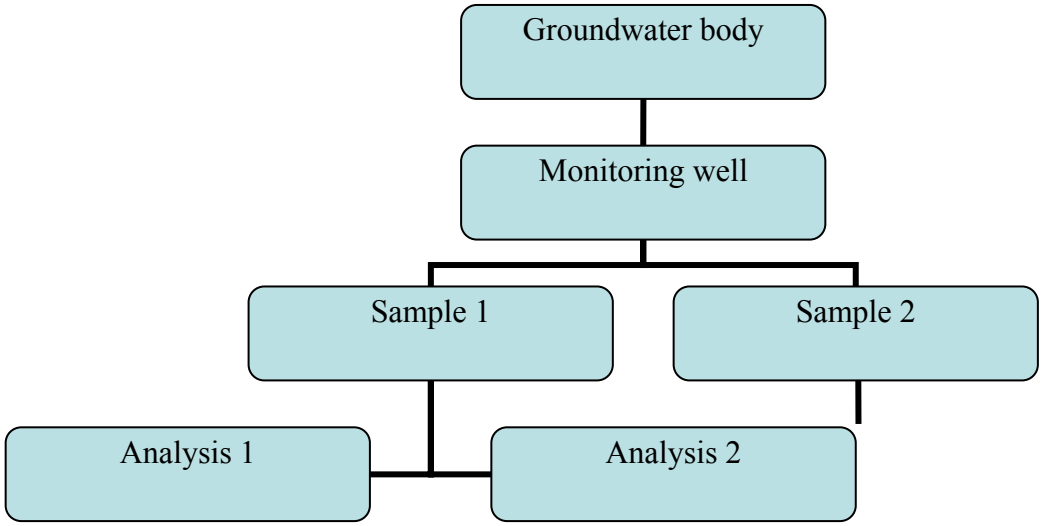


A total of 12 samples were taken and 24 sub-samples were sent for analysis in one sampling round as a validation study.

4.2 Quality control

The objective of the quality control programme for the operational monitoring was to ensure that measurement uncertainty did not increase over time during the monitoring. The quality control programme was set up after careful evaluation of the results from the validation study. Quality control was designed to include duplicate sampling, each with duplicate analysis, on one of the two annual sampling occasions of the monitoring programme, see Figure A3.2. In total, six sampling occasions with 12 samples and 24 sub-samples analysed were included in the first phase of the quality control programme.

Figure A3.2 Design outline for quality control, shown for one sampling occasion



5 Sub-sampling and analysis

The sample pre-treatment and analytical set up for the two key parameters (redox potential and dissolved iron) are shown in Table A3.2.

Table A3.2: Pre-treatment and analytical programme

	Redox potential	Dissolved iron
Pre-treatment	On-line analysed	On-line filtered, preserved with nitric acid, laboratory analysed

5.1 Sub-sampling and sample pre-treatment

Duplicate on-line measurements/sub-samplings for laboratory analysis were done by taking out split sample streams and treating each stream independently. This means that the 'analytical uncertainty' obtained with the duplicate design also included sub-sampling, pre-treatment, such as filtering, and transportation. An estimate of the analytical uncertainty alone could be obtained from the laboratory quality control data, see Section 5.3.

Samples were on-line filtered excluding oxygen through 0.45 µm cellulose acetate membrane filters and sub-samples were preserved in the field for metal analysis by acidification with nitric acid. Sub-samples were stored in polyethylene containers in the dark at less than 10°C during transport to the analytical laboratory.

5.2 Field analysis

The sample stream was pumped through an on-line measuring array of a flow-through cell with sensors set up in series. The sensor used for redox potential is described in Table A3.3.

Table A3.3: On-line sensor used for redox potential measurements

Parameter	Instrument	Cell	Instrument accuracy	Calibration and control
Redox potential	pH 340	Sensolyt Pt	±2 mV	Daily service

No quality control was performed of on-line measurements in the field.

5.3 Laboratory analysis

Analyses were performed at an independent, accredited (ISO 17025) laboratory using accredited methods subject to the required quality assurance and analytical quality control. Methods and performance data from quality control are shown in Table A3.4.

Table A3.4: Methods and performance data from quality control for laboratory analyses

	Method	Within-series repeatability	Between-series reproduci-bility	Total reproducibi-lity	Total expanded uncertainty	Detection limit
Iron	ICP-AES	0.95%	4.2%	4.3%	8.6%	0.01 mg l^{-1}

The certified reference material (CRM) VKI Metal LL2, nominal 0.200 mg Fe l^{-1} was used for quality control with 101.9% recovery (mean for 92 control results).

5.4 Calculation methods

The replicate data were treated using the range method (ISO 3085). For comparison, uncertainty estimates were calculated by analysis of variances (ANOVA) and robust ANOVA (RANOVA) using ROBAN version 1.0.1 (Appendix C3).

The applied calculation methods are demonstrated in Section 7 (below). The range calculations are easily done using standard spreadsheets, and an example can be downloaded from http://team.sp.se/analyskvalitet/sampling/default.aspx.

The occurrence of systematic sampling errors was not assessed quantitatively, but the consistency of the obtained results was used as a qualitative control of systematic errors. As an example, if dissolved iron was found above 0.1 mg l^{-1} in the same sample as oxygen was determined to be above 0.1 mg l^{-1}, this would indicate a systematic sampling and/or pre-treatment error. Similarly, redox potential and oxygen contents were checked to correspond in order to control systematic errors.

6 Results

The data set from the validation study is shown in Table A3.8 for dissolved iron with the range calculations. The calculations for redox potential in the validation study and for both dissolved iron and redox potential during quality control were done similarly.

The data from the validation study (six different wells) using range calculations are shown in Table A3.5.

Table A3.5: Relative expanded uncertainty (%, coverage factor 2) for analysis, sampling and between-target (between wells) as obtained during validation using range calculations

Range calculations	Analyses	Sampling	Between-target
Redox potential	5.2%	15%	14%
Dissolved iron	2.1%	10%	70%

For comparison, the statistical estimates are shown in Table A3.6 as obtained using ANOVA and RANOVA.

Table A3.6: Relative expanded uncertainty (%, coverage factor 2) for analysis, sampling and between-target (between wells) as obtained for dissolved iron during validation using ANOVA and RANOVA calculations

Dissolved iron	Analyses	Sampling	Between-target
ANOVA	1.6%	9.6%	70%
RANOVA	1.8%	9.9%	72%

The statistical estimates obtained with the range statistics during quality control (six sampling occasions) are shown in Table A3.7.

Table A3.7: Relative expanded uncertainty (%, coverage factor 2) for analysis, sampling and between-target (between occasions) as obtained during quality control using range calculations

	Analyses	Sampling	Between-target
Redox potential	18%	3.8%	23%
Dissolved iron	2.5%	3.6%	9.9%

Table A3.8: Results and range calculations for the validation study, dissolved iron, basic data in bold, symbols used to describe calculations only (*T*: target, *S*: sample, *A*: analysis, *R*: absolute differences, *r*: relative differences, *n*: numbers)

Well number	**S1A1** [h]	**S1A2**	**S2A1**	**S2A2**	$R_1 = \lvert S1A1 - S1A2 \rvert$	$\overline{S1} = \frac{S1A1 + S1A2}{2}$	$r_1 = \frac{R_1}{\overline{S1}} * 100$	$R_2 = \lvert S2A1 - S2A2 \rvert$	$\overline{S2} = \frac{S2A1 + S2A2}{2}$	$r_2 = \frac{R_2}{\overline{S2}} * 100$	$\overline{S} = \frac{\overline{S1} + \overline{S2}}{2}$	$r = \frac{\lvert \overline{S1} - \overline{S2} \rvert}{\overline{S}} * 100$
	mg l^{-1}	**mg l^{-1}**	**mg l^{-1}**	**mg l^{-1}**	mg l^{-1}	mg l^{-1}	%	mg l^{-1}	mg l^{-1}	%	mg l^{-1}	%
99.474	**0.815**	**0.834**	**0.912**	**0.893**	0.019	0.825	2.30	0.019	0.903	2.11	0.864	9.03
99.468	**1.80**	**1.83**	**1.94**	**1.93**	0.030	1.82	1.65	0.010	1.94	0.517	1.88	6.40
99.469	**1.69**	**1.68**	**1.79**	**1.77**	0.010	1.69	0.593	0.020	1.78	1.12	1.73	5.48
99.916	**2.62**	**2.61**	**2.83**	**2.84**	0.010	2.62	0.382	0.010	2.84	0.353	2.73	8.07
99.327	**1.66**	**1.63**	**1.58**	**1.59**	0.030	1.65	1.82	0.010	1.59	0.631	1.62	3.72
99.371	**1.52**	**1.53**	**1.47**	**1.50**	0.010	1.53	0.656	0.030	1.49	2.02	1.51	2.66
							$\sum r_1 = 7.413$			$\sum r_2 = 6.750$	$\sum \overline{S} = 10.32$	$\sum r = 35.36$
							$n_{r1} = 6$			$n_{r2} = 6$	$n_r = 6$	$n_r = 6$

Analysis	$r_A = \frac{\sum r_1 + \sum r_2}{n_{r1} + n_{r2}}$	$r_A = \frac{7.413 + 6.750}{6 + 6} = 1.18$	$CV_A = \frac{r_A}{1.128}$ [i]	$CV_A = \frac{1.18}{1.128} = 1.05$	
Sampling	$r_{S+A} = \frac{\sum r}{n_r}$	$r_{S+A} = \frac{35.36}{6} = 5.89$	$CV_{S+A} = \frac{r_{S+A}}{1.128}$	$CV_{S+A} = \frac{5.89}{1.128} = 5.22$	$CV_S = \sqrt{CV_{S+A}{}^2 - \frac{CV_A{}^2}{2}}$ [j] $\quad CV_S = \sqrt{5.22^2 - \frac{1.05^2}{2}} = 5.17$
Between-target	$S_{T+S+A} = \frac{\sum \overline{S}}{n_r}$	$S_{T+S+A} = \frac{10.32}{6} = 1.72$	$s_{T+S+A} = s_{\overline{S}}$ [k]	$s_{T+S+A} = 0.604$	$CV_{T+S+A} = \frac{s_{T+S+A}}{S_{T+S+A}} * 100$ $\quad CV_{T+S+A} = \frac{0.604}{1.72} * 100 = 35.1$
					$CV_T = \sqrt{CV_{T+S+A}{}^2 - \frac{CV_{S+A}{}^2}{2}}$ $\quad CV_S = \sqrt{35.1^2 - \frac{5.17^2}{2}} = 34.9$

[h] S1A1: sample 1 analysis 1.

[i] The standard deviation can be obtained from the mean of relative differences between duplicate measurements by division with the statistical factor 1.128.

[j] The sum of relative variances is $CV_{S+A}{}^2 = CV_S{}^2 + \frac{CV_A{}^2}{2}$ with the factor ½ on $CV_A{}^2$ due to the mean of duplicate analyses being used.

[k] *s*: standard deviation with *n*-1 degrees of freedom as obtained from most standard calculators and spreadsheets.

No groundwater samples had measurements of dissolved oxygen and dissolved iron above 0.1 mg l^{-1}, and the low redox potential measured (-110 to -200 mV) is consistent with the absence of oxygen (<0.1 mg l^{-1}) and the high dissolved iron concentrations (0.92 to 2.8 mg l^{-1}).

7 Comments

Overall, the validation data show that the variability in the aquifer (between-target) was dominating the total uncertainty for dissolved iron, whereas sampling and between-target uncertainties were of the same size for dissolved iron. Analytical uncertainties were small (2–5%), and for dissolved iron were comparable to the repeatability obtained in laboratory quality control (expanded uncertainty 2.1% as compared to 1.9% respectively). If different wells were sampled, the sampling uncertainty was 10–15%.

For dissolved iron measured during validation, the use of ANOVA and RANOVA calculations did not provide statistical estimates more than slightly different from those obtained with the simple range calculations.

In the quality control scheme of monitoring, the variability between sampling occasions (between-target, 9.9%) was dominating the total uncertainty for parameters analysed as laboratory analysis (dissolved iron, 2.5% uncertainty), whereas the analytical uncertainty (18%) was almost as important as the between-target uncertainty (23%) for on-line measurements (redox potential). The reason for the large contribution from on-line measurements is that during quality control, duplicate on-line measurements were done with two different instruments in contrast to the validation study done with one single instrument for both duplicate measurements. Accordingly, the analytical uncertainty (instrument to instrument variation) for redox potential was considerably larger in the quality control (18%) than in the validation study (5.2%). For dissolved iron, the analytical uncertainty was comparable in validation and in the subsequent quality control (2.1% and 2.5% respectively). The sampling uncertainty was lower when sampling just one well on different occasions during quality control (3.6–3.8%) than when sampling different wells at the same time during validation (10–15%). The uncertainty between-target (variation from one sampling occasion to the next) during quality control was small for dissolved iron (9.9%), but larger for redox potential (23%).

If a continuous control of sampling uncertainty had been required, the control data could have been plotted in control charts in order to obtain an early warning of excessive uncertainty (random errors) for each sampling occasion.

The number of replicates (six) in this study was less than used in most cases and the risk of a decreased confidence in the uncertainty estimates should be considered in evaluation of the results.

The uncertainty contribution from sampling bias was only addressed through evaluation of the consistency of the measurements obtained from different, interrelated chemical parameters (oxygen, dissolved iron, redox) and the evaluation supported the conclusion that sampling and sample pre-treatment had succeeded in avoiding bias due to oxygen impact and filter clogging.

8 Summary

The measurement uncertainty (% uncertainty with coverage factor 2) is summarised below for dissolved iron.

The data show that the requirement for less than 20% expanded measurement uncertainty could be fulfilled for dissolved iron (sampling validation), and that the required measurement uncertainty was in reality achieved during the routine monitoring (sampling quality control). Furthermore, the data show that if an improvement of the certainty of monitoring was required, the obvious point of improvement would be increased monitoring density for dissolved iron (between-target uncertainty dominating), whereas improvement of the on-line measurement uncertainty could help for redox potential (large contribution of analysis uncertainty).

Dissolved iron in groundwater	**Expanded uncertainty, coverage factor of 2**			**Between-target variability**
	Sampling	**Analysis**	**Measurement**	
Validation	10%	2.1%	10%	35%[l]
Quality control	3.6%	2.5%	4.4%	9.9%[m]

9 Acknowledgements

The work presented here has been supported by Nordic Innovation Centre, the Soil and Ground Water Contamination Committee of the Danish Council of Technical Sciences and Århus County, Denmark. Fieldwork has skilfully been done by Mogens Wium, GEO.

[l] In the validation study, between-target variability was between wells.

[m] In the quality control, between-target variability was between sampling occasions.

Example A4: Vitamin A in baby porridge containing fruit and milled cereals

Measurand				Uncertainty estimation		
Analyte/ Technique	Unit	Sector/ Matrix	Sampling target	Purpose	Design	Statistics
Vitamin A (as retinol)/ HPLC	µg 100 g^{-1} in powder	Food/ Baby porridge-powder containing fruit	Produced batch	Total measurement uncertainty	Empirical duplicate method	One-way ANOVA

1 Scope

The scope is to estimate the measurement uncertainty and contributions from sampling and analyses. The estimates are based on samples from one type of baby porridge – taken from 10 different batches – using a sampling protocol collecting duplicate samples from each batch.

2 Scenario and sampling target

In the production of baby (infant) porridge, the vitamin A (retinol) is added as a premix (together with vitamin D and vitamin C). The premix is a minor ingredient. All ingredients are mixed thoroughly before distribution into packages. Earlier analysis indicated a bigger variation in analytical result between packages than expected. A measurement uncertainty of 20–30% would be considered acceptable. The question was raised whether the variation is due mainly to analytical uncertainty or to sampling uncertainty. One of the theories suggests that the vitamin is locally unevenly distributed within the package, and therefore will give bigger analytical uncertainty if the test portion is too small (e.g. 3–5 g).[n] One possible explanation of the heterogeneity is that the vitamin premix aggregates in small hot spots, due to electrostatic interactions with the fruit particles in the porridge powder. The producers recommend a test portion size of 40–50 g whenever analysing vitamin A, D and C in baby porridge powder.

In order to compare the measured vitamin A concentration against declared values and European regulatory thresholds, an estimation of measurement uncertainty is desirable. To determine the random component of the measurement uncertainty, an empirical approach using the duplicate method (see Section 9.4.2) was chosen. To estimate the systematic component a comparison with a reference value was made.

[n] EN-12823-1 'Foodstuffs – determination of vitamin A by HPLC' indicates a test sample of approximately 2–10 g.

3 Sampling protocol

Normally a spot sampling approach is employed in which one sample (one package) of a batch is used as a screening sample by comparing its content against the declared values and legal limits.
Validation – In this study two samples are collected from each of 10 different batches of *one type* of baby porridge powder (i.e. 10 sampling targets). Each sample is equal to one package of approximately 400 g powder.
Quality control – Quality control (QC) of sampling from different types of baby porridge is done by collecting two samples from each of eight batches of *different* types of baby porridges (i.e. eight sampling targets). All the types of porridges contain fruit in addition to milled cereals.

To ensure the quality in each package of the product at the time of the 'best before date' of the porridge powder, the producer wraps the product in an air-tight and light-protecting bag. It is therefore assumed the degradation of the vitamin A is negligible during normal shelf life. The sampling for the validation was performed at the place of production. For QC, the samples were purchased partly at the place of production, and partly at the retailers. When the samples were collected from retailers, care was taken to collect the two samples (of each product) at different retailers but in addition to assure the samples had the same batch marking. This is important to avoid adding between-batch variations to the apparent sampling uncertainty, as the sampling protocol in this case specifies sampling from a particular batch.

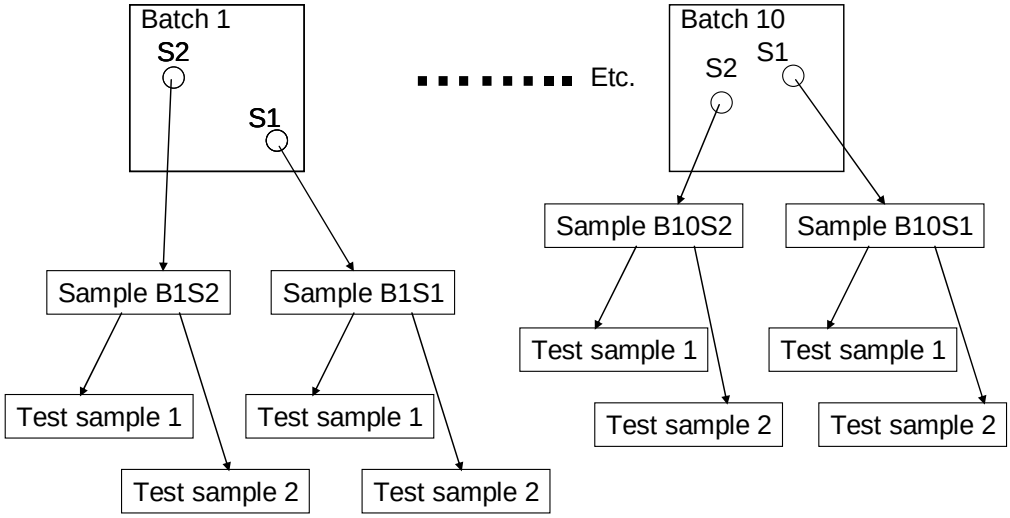
4 Study design – empirical approach

An empirical ('top-down') approach – duplicate method was selected to provide estimates of the random component of sampling uncertainty. The validation is performed on one type of baby porridge containing fruit and milled cereals. In the sampling for the QC different products of baby porridge (all containing fruit and milled cereals) are tested to see if the estimate for measurement uncertainty from the validation study is appropriate for different types of baby porridges containing fruit and milled cereals.

4.1 Validation

Samples are collected on-line (just after the filling operation of packages) at random times. Two samples (two packages, each of approximately 400 g) are collected from each of 10 production units (batches) of one type of baby porridge powder.

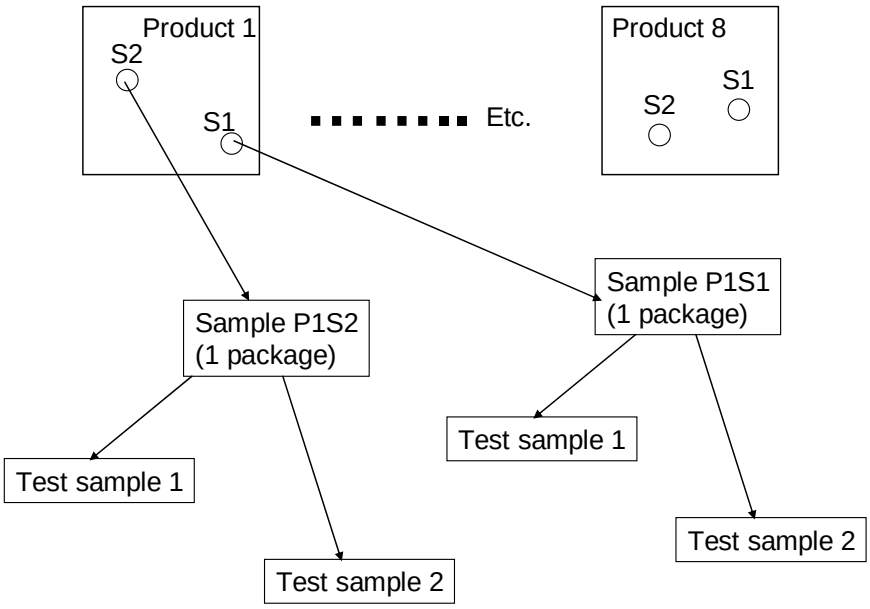
Figure A4.1: Sampling for validation. Two samples are taken from each of 10 production units/batches of the *same type* of baby porridge



4.2 Quality control

For quality control (QC) two samples are collected from one batch of each of eight different types of baby porridges, containing fruit and milled cereals. The porridges are products from three different producers. The samples (except for two types of porridges) were provided by two of the producers. The rest was bought at the retailer.

Figure A4. 2: Sampling for QC. Two samples are taken from one batch of each of eight *different types* of baby porridge



5 Sampling and analysis in the laboratory

The analytical work is done by 'The National Institute of Nutrition and Seafood Research' (NIFES). The laboratory is accredited according to EN ISO/IEC 17025.

The laboratory participates in Laboratory Proficiency Tests (FAPAS and Bipea) with good results (in the period 2000–2005, |Z-score|<1). The method was validated using a certified reference material (CRM). Data concerning the laboratory performance is given in Table A4.1.

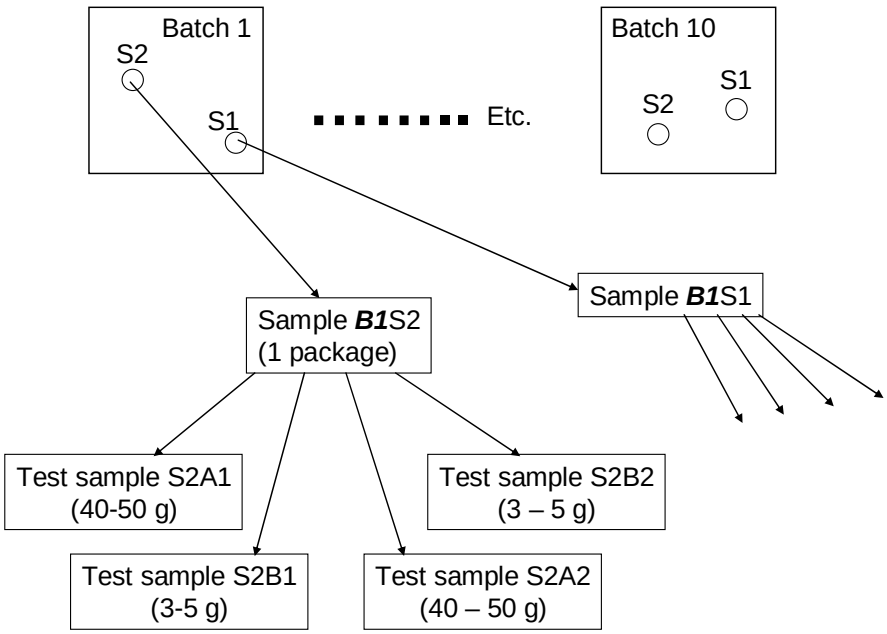
Table A4.1: Methods and performance data from quality control – laboratory analyses

Parameters	Vitamin A – determined as retinol
Method	HPLC – normal phase column – UV-detection
Repeatability	2RSD (%) = 6
Within-reproducibility	2RSD(%) = 8
Measurement uncertainty	14% (95% confidence interval)
Recovery	➢ Standard addition, in lab: 90–110% ➢ Based on laboratory proficiency tests (in period 1999–2005), different matrixes: 88–113%, mean recovery 100.5%
Limit of quantification (LOQ)	0.14 mg kg^{-1}
CRM used	NIST 2383 – baby food (mixed food composite)
➢ CRM – certified level	0.80 ±0.15 mg kg^{-1} (95% confidence interval)
➢ CRM – analysed value	0.77 ±0.14 mg kg^{-1} (*n*=28, 95% confidence interval)

5.1 Secondary sampling

A mechanical sample divider (Retsch) is used to split the samples. From each of the primary samples, four test samples are collected: two portions of approximately 3–5 g and two portions of approximately 40–50 g.

Figure A4.3: Splitting of the primary sample to make four test samples



5.2 Analyses

The analytical method is based on EN-12823-1 (Foodstuffs – Determination of vitamin A by HPLC – Part 1: Measurement of all-trans-retinol and 13-cis-retinol). Retinol is saponified by using ethanolic potassium hydroxide containing antioxidants. Vitamin A is extracted by using hexane. Analysis is performed by using high performance liquid chromatography (HPLC), with a UV detector.

In the validation, for each of the primary samples, two analyses are performed on test samples of 40–50 g and two analyses on test samples of 3–5 g. In the QC two analyses are performed on test samples of 40–50 g. On each test sample only one analytical determination is performed (no analytical duplicates).

6 Information from the producer

Data for estimating the 'true value' of vitamin A in baby porridge are provided by the producer (Nestlé) of the product chosen for the validation, see Table A4.2.

Table A4.2: Product data provided by the producer

Product	Oatmeal porridge with bananas and apricots (Nestlé)
Weight of batch, including premix (1 batch = 2 mixing containers)	1092 kg
Weight of added vitamin-premix in batch	1.228 kg
Vitamin A in premix (data from the Certificate of Analysis)	9016 IU g^{-1} = 2705 µg g^{-1} (retinol)
Vitamin A added to the batch	304 µg 100 g^{-1} (retinol)
Vitamin A in ingredients according to the product specification	45 µg 100 g^{-1} (retinol)
Estimated 'true value' of vitamin A	349 µg 100 g^{-1} (retinol)
Vitamin A declared as	Retinol – (sum of trans- and cis-retinol)

7 Results

Test sample 40 g – baby porridge

Table A4.3: Validation data – from the same product, results given in µg 100 g^{-1} powder

Batch	S1A1	S1A2	S2A1	S2A2
B1	402	325	361	351
B2	382	319	349	362
B3	332	291	397	348
B4	280	278	358	321
B5	370	409	378	460
B6	344	318	381	392
B7	297	333	341	315
B8	336	320	292	306
B9	372	353	332	337
B10	407	361	322	382

S1 and S2: Primary samples from sampling location 1 and 2 of one production batch

A1 and A2: Analyses of duplicate test samples of a primary sample S

Analysed mean value (test sample 40 g): 348 µg 100 g^{-1}

Test sample 4 g – baby porridge

Table A4.4: Validation data – same product, results given in µg 100 g^{-1} powder

Batch	S1B1	S1B2	S2B1	S2B2
B1	400	491	323	355
B2	413	159	392	434
B3	315	391	252	454
B4	223	220	357	469
B5	462	343	262	293
B6	353	265	305	456
B7	298	234	152	323
B8	425	263	417	353
B9	622	189	291	272
B10	292	397	142	568

S1 and S2: Primary samples from sampling location 1 and 2 of one production batch

B1 and B2: Analyses of duplicate test samples of a primary sample S

Analysed mean value (test sample 4 g): 341 µg 100 g^{-1}

7.1 Calculations

The ANOVA calculation can be done by using available tools in Excel, Minitab, SPSS etc. In this study the calculations are done in an excel spreadsheet and the details are shown in Section 10 – ANOVA calculations.

Calculation of uncertainty of analyses, one-way ANOVA, test sample 40 g

Table A4.5: Results from ANOVA calculations – uncertainty of analyses – sum of squares of differences, within-groups (SS-Error). For details see Section A4.11

$SS_{E\text{-}Anal}$ (µg 100 g^{-1})^2	Degrees of freedom (df)	Variance $= SS_{E\text{-}Anal}/df$ (µg 100 g^{-1})^2	Standard deviation, SD_{anal} $=\sqrt{(SS_{EAnal)}/df}$ (µg 100 g^{-1})	Relative standard deviation RSD_{anal}(%) $= (SD / \overline{X}_a)*100\%$
16595	20	829.75	28.805	8.28

Calculation of uncertainty of sampling, one-way ANOVA, test sample 4 g

Table A4.6: Results from ANOVA calculations – uncertainty of sampling – sum of squares of differences. For details see Section A4.11

SS_S (µg 100 g^{-1})^2	Degrees of freedom (df)	Variance $V_{Samp}=$ $(SS_S/df_S - SSE_{Anal}/df_A)/2$ (µg 100 g^{-1})^2	Standard deviation, SD_{samp} $=\sqrt{V_{Samp}}$ (µg 100 g^{-1})	Relative standard deviation RSD_{samp}(%) $= (SD / \overline{X}_s)*100\%$
14231	10	296.7	17.22	4.95

Calculation of measurement uncertainty – 40 g test sample

The RSD (%) value from the ANOVA calculation can be used as an estimate of the standard uncertainty *u* (%). The analytical laboratory has estimated the analytical standard uncertainty to be 7%, which is lower than the random analytical component for this sample type of 8.28%. The higher value of these two is used in the calculations. Combining the RSD values from tables A4.5 and A4.6 with Equation 1, the results can be written as in Table A4.7.

$$u_{meas} = \sqrt{(u_{samp})^2 + (u_{anal})^2} \qquad \text{(Equation A1)}$$

Table A4.7: Measurement uncertainty – 40 g test sample

Measurement uncertainty, ANOVA calculations – 40 g test samples			
	Sampling	Analytical	Total
Uncertainty *u* (%)	4.95	8.28	9.7
Expanded uncertainty *U* (%) = 2**u* With a coverage factor of 2 (i.e. 95% confidence)	9.9	17	20

Calculation of uncertainty of analyses, one-way ANOVA, test sample 4g

The same calculations are used as for the test sample size of 40 g (see Table A4.14, in Section 11 of this example).

Table A4.8: Results from ANOVA calculations – uncertainty of analyses, 4 g test sample – sum of squares of differences, within groups (SS-Error)

SS_E ($\mu g\ 100\ g^{-1})^2$	Degrees of freedom (df) (*N**2-*N*)=20	Variance $= SS_E/df$ ($\mu g\ 100\ g^{-1})^2$	Standard deviation, SD_{anal} $= \sqrt{SS_E/df}$ ($\mu g\ 100\ g^{-1}$)	Relative standard deviation RSD_{anal}(%) = (SD / $\overline{X}_a$)*100%
312206.5	20	15610.325	*124.9413*	36.6800

Calculation of uncertainty of sampling, one-way ANOVA, test sample 4 g

Table A4.9: Results from ANOVA calculations – uncertainty of sampling, 4 g test sample – sum of squares of differences

SS_S ($\mu g\ 100\ g^{-1})^2$	Degrees of freedom (df)	Variance V_{Samp}= ($SS_S/df_S - SSE_{Anal}/df_A)/2$ ($\mu g\ 100\ g^{-1})^2$	Standard deviation, SD_{samp} $= \sqrt{V_{Samp}}$ ($\mu g\ 100\ g^{-1}$)	Relative standard deviation RSD_{samp}(%) = (SD / $\overline{X}_s$)*100%
102860.25	10	-2662.15	$\sqrt{-2662.15}$ Set to zero	Conventionally set to zero

The same calculations are used as for the test sample size of 40 g (Table A4.15, Section 11 of this example).

The negative value of V_{Samp} in Table A4.9 indicates that SD_{Samp} is small compared to the calculated value of SD_{anal}. In this case, the estimates of SD_{anal} and SD_{samp} using robust ANOVA confirmed the smaller sampling standard deviation; the robust ANOVA estimates were $u_{Samp}(\%) = 6.9\%$ and $u_{Anal}(\%) = 30\%$.

As the sampling is identical for the experiments with 40 g and 4 g test samples (and the uncertainty of sampling therefore should be the same), an $RSD_{samp}(\%) = 5\%$ (≈ 4.95 see table A4.7) is used as an estimate.

Calculation of measurement uncertainty – 4 g test sample

Using the calculated RSD (%) value in Tables A4.5 and A4.6 as an estimate of the measurement uncertainty and combining with Equation A1 the results can be written as follows (Table A4.10).

Table A4.10: Measurement uncertainty – 4 g test sample

Measurement uncertainty, ANOVA calculations – 4 g test samples			
	*Sampling	Analytical	Measurement
Uncertainty u (%)	5	36.7	37
Expanded uncertainty U (%) = 2*u With a coverage factor of 2 (i.e. 95% confidence)	10	73.4	74

***** The u (%) value is derived from calculations using 40 g test samples

7.2 Effect of the size of test sample on measurement uncertainty

The baby porridge powder looks homogeneous, and therefore a low measurement uncertainty (u) is expected. However, analyses of the powder indicated a surprisingly large u when using a test sample size of 4 g (the CEN-standard EN-12823-1 indicates a test sample size of approximately 2–10 g). The producers recommended using a test sample size of 40–50 g.

The validation tests gave the following results

Table A4.11: Comparing measurement uncertainty – test samples of 40 g and 4 g

Test sample size	Measurement uncertainty (u_{meas})	Expanded measurements uncertainty U_{meas}
40 g test sample	9.7%	20%
4 g test sample	37%	74%

It can be concluded that $u_{40g} << u_{4g}$. A U_{meas} of approximately 20% is acceptable, using the manufacturer's criterion of 20–30%, while a U_{meas} of 74% is considered to be too high, taking into account the matrix and production conditions of this type of product.

It can therefore be concluded that a test sample weight of 4 g is not 'fit for purpose' when analysing vitamin A (retinol) in baby porridge powder containing milled cereals and fruit. A test sample size of 40–50 g is recommended. This also supports the theory that the vitamin is unevenly distributed in the product, possible as local 'hot spots' due to electrostatic interactions.

7.3 Quality control

According to Section 13.2.2 of this Guide, the principal tool in quality control is replication. This is minimally executed by taking two samples from each target by a complete (and suitably randomised) duplication of the sampling protocol. There is only a need to analyse the sample once and the difference between the results $D = |x_1 - x_2|$ is calculated. In this study each sample was analysed twice, but the comparisons were made between one analyses of each sample (double set).

In the quality control study, test portions of 40 g were used. According to declarations, the products contain different amounts of vitamin A.

Table A4.12: Quality control data for test portion 40 g of different products

Product	Producer	Porridge powder ingredients	S1A1	S1A2	S2A1	S2A2
P1	1	Oat, rice and pear	322	319	350	375
P2	1	Oat, rye, rice and pear	332	317	358	393
P3	1	Wheat, banana and apple	443	430	461	388
P4	1	Wheat and apple	318	383	390	334
P5	2	Oat, rice and banana	252	219	265	227
P6	2	Wheat and apple	274	239	233	217
P7	2	Oat, rice and apple	206	225	198	195
P8	3	Wheat, spelt, oat and apple (organic product)	392	335	375	416

S1 and S2: Primary samples (laboratory samples) from sampling locations 1 and 2 of one batch from each product.

A1 and A2: Analyses on two test samples from each laboratory sample.

Quality control – calculation and control chart

The validated uncertainties of sampling and analysis are u_{samp} and u_{anal} respectively. The construction of a control chart is described in Section 13.2. In the case of baby porridge (40 g test sample) the following calculations can be made:

Warning limit**:** $WL = 2.38 * \sqrt{u^2_{anal} + u^2_{samp}} = 2.83 * \sqrt{\left(4.95^2 + 8.28^2\right)}\% = 27\%$

Action limit**:** $AL = 3.69 * \sqrt{\left(4.95^2 + 8.28^2\right)}\% = 36\%$

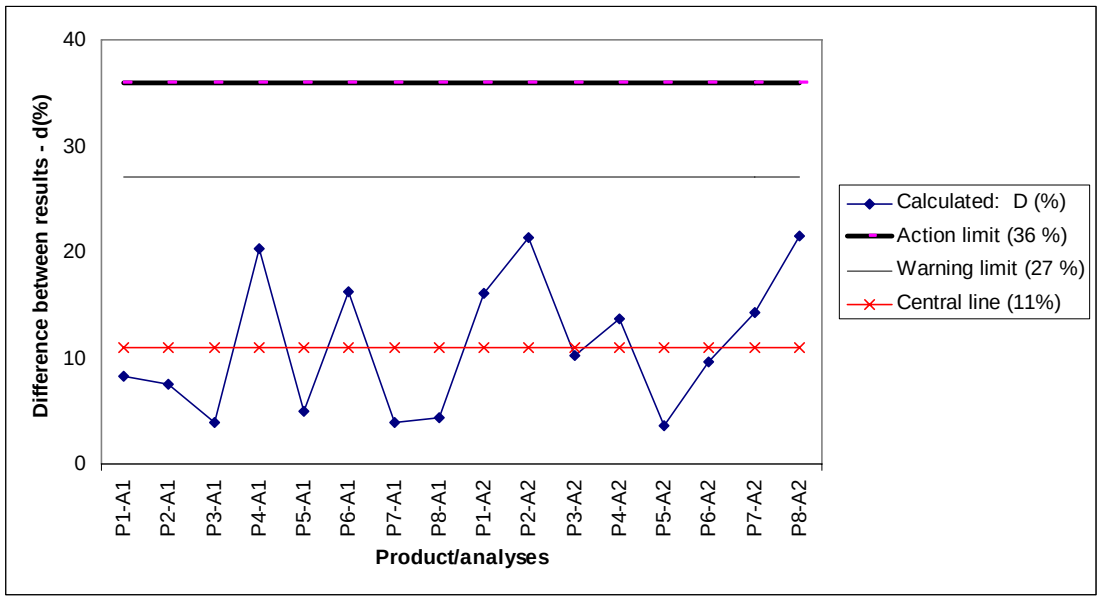
Central line: $CL = 1.128 * \sqrt{\left(4.95^2 + 8.28^2\right)}\% = 11\%$

Table A4.13: Quality control: calculation of differences *D* and *D* (%) – between samples from a batch

Product	Analyses	Sample S1 X_{S1}	Sample S2 X_{S2}	$D = \lvert x_{S1} - x_{S2} \rvert$	$\bar{x}$	$D(\%) = (D/\bar{x}) * 100\%$
P1	A1	322	350	28	336	8
P2		332	358	26	345	8
P3		443	461	18	452	4
P4		318	390	72	354	20
P5		252	265	13	259	5
P6		274	233	41	254	16
P7		206	198	8	202	4
P8		392	375	17	384	4
P1	A2	319	375	56	347	16
P2		317	393	76	355	21
P3		430	388	42	409	10
P4		383	334	49	359	14
P5		219	227	8	223	4
P6		239	217	22	228	10
P7		225	195	30	210	14
P8		335	416	81	376	22

The calculated *D* (%) in Table A4.13 can be compared directly with the action limit, or the results can be presented in a control chart, see Figure A4.4.

Figure A4.4: Control chart: quality control analyses of vitamin A in baby porridge containing cereals and fruits



The control chart in Figure A4.4 shows that when collecting duplicated samples from the same batch, the difference between analytical results *D* (%) is smaller than the action limit. All the calculated difference are in fact smaller than the warning limit of 27%.

The measurement uncertainty determined in the validation step is therefore considered suitable for the quality control of the sampling of baby porridge containing milled cereals and fruit.

If the normal procedure is to analyse one sample from each batch, it is recommended that duplicate samples are collected from the same batch at least in one of ten of the sampled batches.

Measurement uncertainty

Sampling uncertainty

Calculations from the validation study gave an expanded sampling uncertainty U_{samp} (%) = 10% (40 g test sample – see Table A4.7). The calculated uncertainty does not include contributions to the uncertainty due to 'between protocol' and 'between samplers' differences.

Analytical uncertainty

Calculation from the validation study gave an expanded measurement uncertainty of analyses (U_{anal}) of 17% – 40 g test sample. The laboratory reports their own estimation of the analytical uncertainty (see Table A4.1): $2*RSD_{inlab}(\%)$ = 14%. $2*RSD_{inlab}(\%)$ is used as an estimate of U_{anal} in the laboratory. The U_{anal} found in the validation study was at the same level but still a little bigger than the Ua_{nal} reported by the laboratory.

The certified reference material (CRM) used is 2383 (NIST) – baby food composite. The CRM is a mix of different foods of plant and animal origins – and the uncertainty found when analysing the CRM might not be identical with that found when analysing baby porridge powder. Laboratory data for the CRM 2383 is included in the table below.

CRM 2383	Mean value mg kg^{-1}	U $(\%)_{95\%}$	Laboratory bias (%)
Certified	0.80 ± 0.15	18.8	-
Analysed	0.77 ± 0.14	18.2	- 3.75

The measurement uncertainty and the bias determined for the CRM could be allowed for in the analytical measurement uncertainty (as in the NordTest UFS Guide, Example 2), but as the matrix in the validation study is different from that for the CRM used, we chose not to include it in this study.

Total measurement uncertainty

Calculations from the validation study gave an expanded measurement uncertainty $U_{meas}(\%) = 20\%$ (40 g test sample – see Table A4.7).

Systematic bias

The laboratory reports a recovery of normally 90–110%. Recovery based on laboratory proficiency tests 1999–2005: 88–113%. The results for the PT indicate no (or very small) systematic bias. Analyses of CRM 2383 in the laboratory gives a mean analysed value of 96.3% of the certified value – which indicates a small bias (-3.7%). As the matrix of the CRM 'baby food composite' is different to the baby porridge, and the analytical method includes an extraction, the bias determined when analysing the CRM might not be representative for the analyses of baby porridge.

In the validation study, the mean value of retinol was determined to be 348 µg 100 g^{-1} (when using a test sample of 40 g). According to data provided by the producer (see Table A4.2), the 'true value' for retinol was calculated to be 349 µg 100 g^{-1} porridge powder. This gives a recovery of 99.7% of the 'true value'. This indicates that the systematic error due to sampling and analyses is small and might be negligible when analysing baby porridge-powder containing milled cereals and fruits – on the condition that a test sample of 40–50 g is used.

8 Comments

When a test sample of approximately 40 g is used, the retinol concentration C in baby porridge-powder containing milled cereals and fruit should be reported with the expanded measurement uncertainty, i.e. C±20% of the measured value C (95% confidence).

When baby porridge-powder containing milled cereals and fruit is to be analysed, it is recommended to use a relatively large test sample of approximately 40–50 g and not 2–10 g as indicated in the official CEN method (EN-12823-1). As the analytical uncertainty (40 g test sample) was bigger than the normal analytical uncertainty of the laboratory, even larger samples than 40 g might be considered.

9 Summary

Measurement uncertainty for **40 g** test samples				Sample
	Sampling	Analytical	Total	Typical between-target variation RSD_B(%) of the mean values of analyses of the batches in the validation test (see Table A4.15)
Uncertainty *u* (%) = RSD (%)	4.95	8.3	9.7	8.2
[#]Expanded uncertainty *U* (%) = 2**u*	9.9	16.6	19.4	16.4

[#] With a coverage factor of 2 (i.e. 95% confidence)

10 Acknowledgements

Nestlé (Norway) is thanked for its enthusiastic cooperation and in addition for providing samples to the project (validation and quality control study). Also Smaafolk – Tine Norske Meierier is thanked for kindly offering us samples for the quality control study. The National Institute of Nutrition and Seafood Research (NIFES) is thanked for the analytical contribution (analyses and information on the laboratory QA-system). The study was done with financial support from the Nordic Innovation Centre and the Norwegian Food Safety Authority.

11 ANOVA calculation, vitamin A in baby porridge – details

Calculation of uncertainty of analyses, one-way ANOVA, test sample 40 g

Table A4.14: ANOVA calculations – uncertainty of analyses – sum of squares of differences, within groups (SS-Error)

Sample	Analyses (µg 100 g^{-1})		Mean value – each sample (µg 100 g^{-1})	Squares of differences – within groups (µg 100 g^{-1})^2
	A1 =$x_{i\,j}$ =x_{i1}	A2 = $x_{i\,j}$ = x_{i2}	$\overline{x_i}$ = (x_{i1}+ x_{i2})/2	(x_i- $\overline{x}_i$)^2
B1-S1	402	325	363.5	1482.25
B2-S1	382	319	350.5	992.25
B3-S1	332	291	311.5	420.25
B4-S1	280	278	279	1
B5-S1	370	409	389.5	380.25
B6-S1	344	318	331	169
B7-S1	297	333	315	324
B8-S1	336	320	328	64
B9-S1	372	353	362.5	90.25
B10-S1	407	361	384	529
B1-S2	361	351	356	25
B2-S2	349	362	355.5	42.25
B3-S2	397	348	372.5	600.25
B4-S2	358	321	339.5	342.25
B5-S2	378	460	419	1681
B6-S2	381	392	386.5	30.25
B7-S2	341	315	328	169
B8-S2	292	306	299	49
B9-S2	332	337	334.5	6.25
B10-S2	322	382	352	900
Mean value of measurements: $\overline{X}_a$ =1/20 ∗ $\sum_{i=1}^{20}$ $\overline{x}_i$ = 347.85 µg 100 g^{-1}			[2] **SS-Error** (SS_E): = $\sum_{i=1}^{20}$ $[(x_{i1}- \overline{x}_i)^2+(x_{i2}- \overline{x}_i)^2]$= $\sum_{i=1}^{20}$ 2_* $(x_i- \overline{x}_i)^2$	
SS_E (µg 100 g^{-1})^2	Degrees of freedom (df) (N*2-N)=20	Variance = SS_E /df (µg 100 g^{-1})^2	Standard deviation, SD_{anal} = $\sqrt{SS_E/df}$ (µg 100 g^{-1})	Relative standard deviation RSD_{anal}(%) = (SD / $\overline{X}_a$)*100%	
16595	20	829.75	*28.80538*	8.280978	

Notes on Table A4.14.

1. Calculation of SS-Error – in this case two test samples are analysed for each laboratory sample, therefore:

$$(x_{i1}-\bar{x}_i)^2=(x_{i2}-\bar{x}_i)^2 \Rightarrow SS_E=\sum_{i=1}^{20}\left[(x_{i1}-\bar{x}_i)^2+(x_{i2}-\bar{x}_i)^2\right]=2\sum_{i=1}^{20}(x_{i1}-\bar{x}_i)^2$$

If the number of test samples analysed is greater than two, the squares of differences will be not be equal and the calculation to be done is the following: $SS_E = \sum_{i=1}^{20} \sum_{j=1}^{n} \left(x_{ij} - \bar{x}_i\right)^2$

2. df = (*N***n*-*N*)=(20*2-20)= 20 where *N* is the number of samples and *n* is the number of test samples analysed of each batch.

Calculation of uncertainty of sampling, one-way ANOVA, test sample 40 g

Table A4.15: ANOVA calculations – uncertainty of sampling – sum of squares of differences

$\overline{x_i}$ S1A1=x_{i1}	S1A2=x_{i2}	S2A1=x_{i3}	S2A2=x_{i4}	$\overline{x_i}$	$\left(\frac{x_{i1}+x_{i2}}{2}-\bar{x}_i\right)^2$	$\left(\frac{x_{i3}+x_{i4}}{2}-\bar{x}_i\right)^2$
402	325	361	351	359.75	14.0625	14.0625
382	319	349	362	353	6.25	6.25
332	291	397	348	342	930.25	930.25
280	278	358	321	309.25	915.0625	915.0625
370	409	378	460	404.25	217.5625	217.5625
344	318	381	392	358.75	770.0625	770.0625
297	333	341	315	321.5	42.25	42.25
336	320	292	306	313.5	210.25	210.25
372	353	332	337	348.5	196	196
407	361	322	382	368	256	256

$$SS_{Samp} = \sum_{i=1}^{10}\left[\left(\frac{x_{i1}+x_{i2}}{2}-\bar{x}_i\right)^2+\left(\frac{x_{i1}+x_{i2}}{2}-\bar{x}_i\right)^2+\left(\frac{x_{i3}+x_{i4}}{2}-\bar{x}_i\right)^2+\left(\frac{x_{i3}+x_{i4}}{2}-\bar{x}_i\right)^2\right]$$

$$= \sum_{i=1}^{10}\left[2*\left(\frac{x_{i1}+x_{i2}}{2}-\bar{x}_i\right)^2+2*\left(\frac{x_{i3}+x_{i4}}{2}-\bar{x}_i\right)^2\right] = 14231$$

Mean value of all measurements $\bar{x}$ = 347.85	$RSD_{Samp}(\%)=(SD_{Samp}/\bar{x})*100\%$= 4.95%
SSE_{Anal} = 16595 (see Table A4.14)	df_S =10 (see table note) df_A = 20 (see Table A4.14)
Variance $V_{Samp}= (SS_S/df_S - SS_A/df_A)/2$ = (14231/10 – 16595/20)/2 = 296.675	$SD_{Samp} = \sqrt{V_{samp}}$ = 17.224

Notes on Table A4.15.

1. The difference d between the mean value $\bar{x}$ of the two values $\left(\frac{x_{i1}+x_{i2}}{2}\right)$ and $\left(\frac{x_{i3}+x_{i4}}{2}\right)$ to each of the values are identical. The expression could therefore be written as

$$SS_{Samp} = \sum_{i=1}^{10} 4*d_i^{\,2} = \sum_{i=1}^{10}\left[4*\left(\frac{x_{i1}+x_{i2}}{2}-\bar{x}_i\right)^2\right]$$

2. dfs= (*NB***n*-*NB*)=(10*2-10)= 10 where *NB* is the number of batches and *n* is the number of primary samples (= laboratory samples) analysed for each batch.

Example A5: Enzyme in chicken feed

Measurand				Uncertainty estimation		
Analyte/ Technique	Unit[1]	Sector/ Matrix	Sampling target	Purpose	Design	Statistics
Enzyme/ HPLC	% m/m (i.e. mass fraction)	Food & Feed/ Chicken feed	25 kg bag	Total uncertainty (weak links in measurement chain)	Modelling with sampling theory (of Gy)	Summation of component variances

[1]Including reporting base

1 Scope

The scope is to estimate the sampling uncertainty with the given sampling protocol by applying Gy's sampling theory (Section 10.2). The analyte is an added enzyme ingredient in the feed. Sampling theory provides realistic estimates only if all sampling and sample splitting operations are carried out obeying the rules of sampling correctness; it is assumed in this example that no gross errors are present and that 'incorrect sampling errors' are negligible.

2 Scenario and sampling target

An enzyme product is used as an additive in chicken feed (density = 0.67 g cm^{-3}). The nominal concentration of the enzyme is 0.05% m/m. The enzyme powder has a density of 1.08 g cm^{-3}. Powders are carefully mixed. The size distribution of the enzyme particles was known and it was estimated that the characteristic particle size was d = 1.00 mm and the size factor was g = 0.5. The purpose of this exercise is to estimate the total uncertainty of the protocol (i.e. as fundamental sampling error, Section 10.2.7 and Figure 4) used for estimating the average content in each 25 kg bag employed to ship the product to customers.

3 Study design, using a modelling approach ('bottom-up')

A model is constructed using sampling theory as described in Section 10.2. The parameters are either measured directly, or estimated, and assumed to be single values and to be constant within and between each bag.

4 Sampling and analysis in the laboratory

The actual concentration of the enzyme in the sampling target, which is identified as a 25 kg bag, is estimated by taking a 500 g primary sample from it.

The material from the primary sample is ground to a particle size of <0.5 mm. Then the enzyme is extracted from a 2 g test portion by using a suitable solvent and the concentration is determined by using liquid chromatography. The relative standard deviation of the chromatographic measurement, estimated from the laboratory quality control data, is 5%.

5 Results

To estimate the fundamental sampling error (FSE, Section 10.2.7, Figure 4) of the two sampling steps, we have the to evaluate the material properties (Table A5.1).

Table A5.1: Input values for estimation of sampling uncertainty by the modelling approach, using sampling theory

Primary sample	**Secondary sample**	**Comment**
M_1 = 500 g	M_2 = 2.0 g	Sample sizes
M_{L1}= 25,000 g	M_{L2} = 500 g	Lot (sampling target) sizes
d_1 = 0.1 cm	d_2 = 0.05 cm	Particle sizes
g_1 = 0.5	g_2 = 0.25	Estimated size distribution factors
Both samples		
a_L = 0.05% m/m		Mean concentration of enzyme in the lot
α = 100% m/m		Enzyme concentration in enzyme particles
ρ_c = 1.08 g cm^{-3}		Density of enzyme particles
ρ_m = 0.67 g cm^{-3}		Density of matrix particles
f = 0.5		Default shape factor for spheroidal particles
β = 1		Liberation factor for liberated particles

These material properties give for the constitution factor (Equation 7) the value c = 2160 g cm^{-3} and for the sampling constants (Equation 6) C values

C_1 = 540 g cm^{-3} and C_2 = 270 g cm^{-3}

Equation 5 can be used to give estimates of the standard deviation for each sampling step (as estimates of the standard uncertainty).

s_{r1} = 0.033 = 3.3% …. Primary sample

s_{r2} = 0.13 = 13% …. Secondary sample

s_{r3} = 0.05 = 5% …. Analytical determination

The total relative standard deviation (s_t, combined uncertainty) can now be estimated by applying the rule of propagation of errors; for i errors we have:

$$s_t = \sqrt{\sum s_{ri}^2} = 0.143 = 14.3\ \%$$

The relative expanded uncertainty, with a coverage factor of 2, is therefore 28.6% (excluding analytical uncertainties associated with systematic effects, such as analytical bias).

6 Comments

The largest source of uncertainty in the whole measurement process is identified as that generated in preparing the test portion (2 g) for the extraction of the enzyme.

No additional allowance has been made for uncertainties associated with systematic effects during analysis, and incorrect sampling errors (and sampling bias) have been assumed to be negligible.

7 Assessment of fitness for purpose of these measurements

If it is decided that the overall uncertainty of 28.6% is not fit for purpose (Section 16), then it is the step in which the test portion is prepared that needs to be modified, to reduce the overall uncertainty. Either a larger sample should be used for the extraction, or the primary sample should be pulverised to a finer particle size, whichever is more economic in practice. The model can also be used to predict either the increase in mass, or reduction in particle size, that is required to achieve the uncertainty that will be considered fit for purpose (e.g. Appendix E).

8 Reporting and interpretation

Measurement of the enzyme concentration reported for each 25 kg bag should have an attached uncertainty of 28.6% of the concentration value. The continued use of this uncertainty value will depend on the periodic checking of the validity of the values and assumptions used in its calculation.

9 Summary

Measurement uncertainty*		
Sampling	Analytical	Total
26.8% (rel)	10.0% (rel)	28.6% (rel)

* with coverage factor of 2 (i.e. 95% confidence)

Example A6: Cadmium and phosphorous in agricultural top soil by modelling approach

Measurand				Uncertainty estimation		
Analyte/ Technique	Unit	Sector/ Matrix	Sampling target	Purpose	Design	Statistics
Cd: GF-ZAAS direct solid sampling P: Ca-Acetate Lactate (CAL) method	mg kg^{-1} to air dried basis	Environ-mental/ agricultural top soil	Arable soil – 143 x 22 m, depth 30 cm	Total uncertainty (with contributions from each sampling effect)	Modelling approach (using exploratory measure-ments for single effects)	Summation of component variances

1 Scope

Estimation of the overall uncertainty of measurement by summation of individual uncertainty contributions from sampling, sample preparation and analysis using the modelling approach.

2 Scenario and sampling target

Investigation aims to estimate the mean concentration of cadmium and phosphorus in top soil of a target that is an area of arable land of 0.32 hectare (specification of the measurand). Sampling used composite samples in a protocol that is commonly applied to agricultural control.

3 Sampling protocol

The target area was sampled using a stratified protocol, with a sampling density of approximately 20 increments per hectare, to a depth of 30 cm, using a soil auger.

4 Study design – cause-and-effect modelling approach (Section 10.1)

4.1 Identification of effects in the measurement

The following sources can be considered as potential significant contributors to the uncertainty in the general case.

4.1.1 Sampling

The spatial distribution of the analyte over a two-dimensional object creates two different uncertainty components ‘long range point selection error’ (Appendix C2.3):

- The *sampling variance* of the analyte content between composite samples from different locations characterises the ’statistical distribution’ of the analyte over the target area. This value often depends on the distance between sampling points/sampling locations.

- If the spatial pattern of the analyte on the area is not represented by the sampling pattern (sampling strategy), sampling bias may occur.

With the use of a sampling tool, different effects may appear, such as point materialisation error (Figure 3). This may occur due to an ill-defined reference level of the soil (e.g. due to undulation of the soil surface or difficulties in the definition of the horizons), or variation in the actual sample depth or in soil density (e.g. by moisture content), or by selective loss of soil material from the sampling device.

These effects only lead to an uncertainty contribution, if there is a depth gradient in the analyte content (a 'third dimension' to the target body). For this reason, these effects, which are difficult to determine one by one, are summarised collectively as the *'depth-effect'*.

4.1.2 Sample preparation

The physical sample preparation comprises the step from the field sample to the laboratory sample. *Mechanical treatment*, such as disaggregation, sieving, grinding and splitting steps, reduce the amount of soil material. With these steps errors may arise due to variation in duration and forces of mechanical treatment, heterogeneity, segregation of different soil (particle) fractions and particles size distribution. A periodic point selection error (Figure 3) may occur due to *variation in moisture content* of the dried soil sample by sorption/desorption of water from air to an equilibrium state (depending on the humidity and properties of the sample material, e.g. particle size).

4.1.3 Analysis

The analysis is the third step of the measurement process, which is connected with different kinds of effects that give rise to uncertainty contributions. The analytical uncertainty of the laboratory samples can be estimated by previously published procedures [1, 35]. The separation of the laboratory sample into analytical test samples will add to the sampling uncertainty; specifically, another 'fundamental error' may occur. However, the random component of this sampling effect is included in the analytical repeatability precision between test samples. A significant systematic component should be avoided by proper mixing of the sampling powder.

4.2 Cause-and-effect diagram

Figure A6.1 shows the 'cause-and-effect diagram' for the measurement process. In the sampling and sample preparation steps the sources of uncertainty contributions are given; for the analysis, only the analytical quality parameters are indicated.

4.3 Model equation

The 'input quantities' of the sampling effects discussed above are not constituent parts of the equation from which the measurement result is calculated. An appropriate model equation for the overall measurement process can be established, however, by introducing respective nominal correction factors on the analytical result:

$$x_{site} = \overline{x}_{anly} \times f_{b\text{-}loc} \times f_{strat} \times f_{depth} \times f_{prep} \times f_{dry}$$

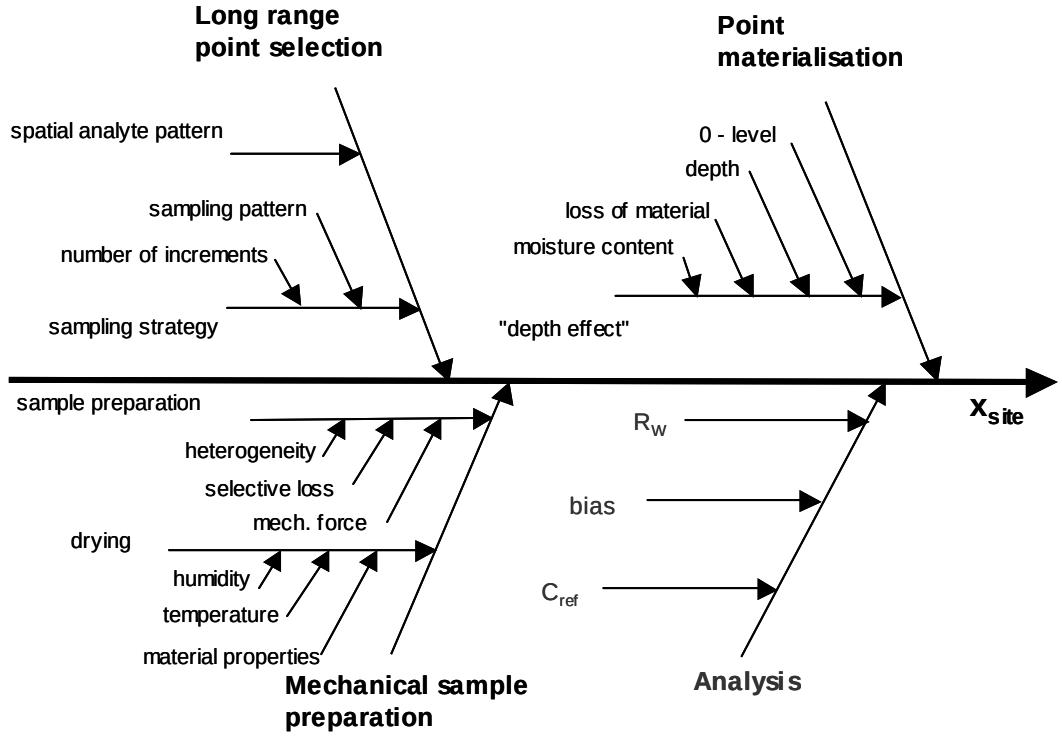
where

x_{site} = measurement result

$\overline{x}_{analy}$ = mean from the analysis of test samples

$f_{b\text{-}loc}$ = correction factor for deviation 'between locations'
f_{strat} = correction factor for bias due to sampling strategy
f_{depth} = correction factor for the 'depth effect'
f_{prep} = correction factor for errors during mechanical sample preparation
f_{dry} = correction factor for deviation of moisture content

Figure A6.1: Cause-and-effect diagram for soil sampling on arable land (R_w is within-laboratory reproducibility)



If no significant bias is detected, all correction factors can be set to unity so that the best estimate for the measurand is given by:

$$x_{site} = \bar{x}_{anly}$$

Because of the simplicity of the model equation (only factors), and assuming independence between the factors, the combined uncertainty can be achieved by variance addition of the *relative standard uncertainties* from the various effects:

$$u_{site} = \sqrt{u_{anly}^2 + u_{b\text{-}loc}^2 + u_{strat}^2 + u_{depth}^2 + u_{prep}^2 + u_{dry}^2}$$

5 Sampling and analysis in the laboratory

The sample mass was reduced by cone and quartering, air dried and sieved to select the grain size <2 mm.

The analysis was performed by the following methods for cadmium using Zeeman-GF-AAS ('direct solid sampling') and for phosphorus using the Calcium-Acetate-Lactate (CAL) method (the analytical measurement procedures are described elsewhere in separate protocols).

6 Results of evaluation of individual effects in this case study

The estimation of the standard uncertainty from the analyte distribution over the target area is based on a modified increment sampling based on the sampling protocol. For elucidation of the outcome of single effects additional exploratory measurements have been carried out.

6.1 Variation 'between locations'

The area was divided into nine squares (A, B, C × 1, 2, 3), and three increments are taken from each of five squares ('crosswise' over the area). The increments from each square are combined, resulting in five separate composite samples. These samples are treated and analysed separately. The mean of the single results constitutes the measurement result in agreement with the specification of the measurand.

The analytical results for both analytes under investigation are shown in Table A6.1. The standard deviation between these values (s_{sqr}) reflects the variation between the composite samples for each nominate square.

The standard uncertainty in the overall mean value (i.e. the measurement result) due to this effect can be estimated by considering the number of samples 'between locations' using the standard error on the mean:

$$u_{b\text{-}loc} = \frac{s_{sqr}}{\sqrt{n_{b\text{-}loc}}}$$

Table A6.1: Measured concentration of cadmium and phosphorus in five squares

Square	Cd mg kg^{-1}	P mg kg^{-1}
A1	0.270	124
A3	0.285	112
B2	0.343	120
C1	0.355	118
C3	0.343	105
$\overline{x}_{sqr}$	0.319	116
s_{sqr}	0.039	8.0
	(12%)	(6.5%)
$u_{b\text{-}loc}$	5.4%	2.9%

The table shows the mean value across the five squares (the measurement result), the standard deviation calculated from these values (s_{sqr}), and the estimated uncertainty contribution from the standard error on the mean ($u_{b\text{-}loc}$).

6.2 Sampling strategy

Inspection of the analyte contents between the squares (Table A6.1) shows no notable differences for phosphorus in any direction (neither vertical, nor horizontal, nor diagonal). So, no significant bias (e.g. ≤ 0.5%) in the measurement result can be expected for this analyte from this source.

For cadmium both A squares show a considerably lower analyte content than the B and C squares. Such a gradient was not unexpected for this particular area because the C squares lay on a forest boundary, while the A squares border on grassland and the 1 and 3 squares lay between other arable land areas. It is well known that in the upper horizon of forest soils accumulation of heavy metal occurs, which can influence adjacent areas.

A 'hypothesis-based' sampling pattern was applied to look for such an effect. However, the values measured with this sampling strategy only detected a minor systematic effect. A standard uncertainty of ≤ 1% is therefore inserted into the uncertainty budget for cadmium for sampling strategy.

6.3 'Depth effect'

For revealing the collection of effects referred to as the 'depth effect', the following exploratory experiment was performed.

Increment cores are taken for a depth of 35 cm within the five 'test squares'. From these cores segments of 25–30 cm and of 30–35 cm are separated and combined. Table A6.2 shows the analytical results for these samples.

Table A6.2: Depth experiments

	Cd mg kg^{-1}	P mg kg^{-1}
c_- (25–30 cm)	0.14	47
c_+ (30–35 cm)	0.10	35
x_-	0.34	124
x_+	0.30	109
Δx	0.04	15
u_{depth}	3.5%	3.7%

The table shows the average content of the depth horizons from five cores from different locations, the calculated content limits and the estimated uncertainty contribution

Both analytes show a statistically significant negative gradient with respect to depth. The uncertainty due to the depth effect was estimated by considering the analyte content of the soil layers below and above the reference depth (c_-, c_+) by the following model.

The maximum variation in the sampling depth is assumed to be not more than ±10% (i.e. 27–33 cm). From these data the lower and upper content limits (x_-, x_+), related to the mean content of an auger core of nominal depth, are estimated according to:

$$x_- = \frac{\bar{x} - 0.1 c_-}{0.9} \qquad x_+ = \frac{\bar{x} + 0.1 c_+}{1.1}$$

The difference between x_- and x_+ (Δx_{depth}) is assumed to be the maximum deviation from the mean content due to depth variation of the increments.

If a rectangular distribution for the deviation in depth is assumed, the standard uncertainty in the mean value (Table A6.2) can be estimated by:

$$u_{depth} = \frac{\Delta x_{depth} / 2}{\sqrt{3}}$$

6.4 Splitting

The primary field samples were split in half, seven times by a coning and quartering procedure resulting in a laboratory sample that was 1/64 of the original mass.

To reveal the 'splitting effect' the following exploratory experiment was performed.

In the first splitting step the second half of the material was not discarded, but considered as a duplicate sample, which was treated like the original sample and analysed separately. Table A6.3 shows the relative standard deviations between the duplicates of each of the five squares for both analytes.

As a simple approximation, the mean of the relative standard deviations is taken as the standard uncertainty of the splitting step

$$u_{split} = \bar{s}_{split}$$

Note: The observed large spread of standard deviations between the duplicates must be expected. The χ^2-distribution for df = 1 shows high probability for very low values and a moderate probability for large values.

Table A6.3: Relative standard deviations between duplicate split samples and the mean of these standard deviations for both analytes

Square	Cd (%)	P (%)
A1	0.44	1.49
A3	9.17	2.80
B2	5.32	0.84
C1	3.24	8.88
C3	0.44	1.81
$\bar{s}_{split}$	3.7	3.3

6.5 Drying

For the drying effect no experiment was performed, but information from the literature was used to estimate the effect. A moisture content between 1 and 3% has been found for a large number of air-dried soil samples [44]. According to the sampling protocol, the measurand refers to air-dried soil material. Consequently, no correction for moisture content is required for the concentration measurements. However, a range of $\Delta x_{dry} = 2\%$ difference in moisture content must be considered. Assuming a rectangular distribution across this range, the standard uncertainty for both analytes can be estimated as:

$$u_{dry} = \frac{\Delta x_{dry} / 2}{\sqrt{3}} = 0.6\%$$

6.6 Analysis

The uncertainty from the analytical process for cadmium and phosphorus (Tables A6.4 and A6.5) were estimated from quality control data, using the Nordtest-approach [25].

Table A6.4: Standard uncertainty components and combined uncertainty in the analysis of the soil sample for cadmium

R_W	Uncertainty from within-laboratory reproducibility, evaluated from the repeatability standard deviation of the mean from n=10 test samples and the instrument stability over the working session of one day	u_{Rw} = 3.6%
c_{ref}	Uncertainty of the certified value of a CRM	u_{ref} = 2.7%
bias	No uncertainty contribution from laboratory bias, because the results are corrected for the day-to-day bias of the CRM measurements	-
s_{bias}	Uncertainty contribution from the standard deviation of the mean (n=3) from the day-to-day analysis of the CRM	u_{bias} = 2.7%
	Combined analytical uncertainty	u_{anly} = 5.2%

Table A6.5: Standard uncertainty components and combined uncertainty in the analysis of the soil sample for phosphorus

R_W	Uncertainty from within-laboratory reproducibility, evaluated from the repeatability standard deviation of the mean from n=1 test samples	u_{Rw}= 1.7%
c_{ref} bias s_{bias}	Uncertainty for the trueness of the results estimated as the reproducibility precision s_R from one inter-laboratory comparison (worse case estimate)	u_{bias} = 9.5%
	Combined analytical uncertainty	u_{anly} = 9.7%

6.7 Uncertainty budget and measurement result

Table A6.6 lists the evaluated standard uncertainty from the effects under consideration. The combined uncertainty is calculated from these contributions.

Table A6.6: Relative standard uncertainties from the considered effects and the combined uncertainty for both analytes

Effect	Relative standard uncertainty (%)	
	Cd	P
Variation 'between locations'	5.4	2.9
Sampling strategy	1.0	0.5
Depth	3.5	3.7
Splitting	3.7	3.3
Drying	0.6	0.6
Analysis	5.2	9.7
Combined uncertainty	**9.1**	**11.3**

Measurement result:

Cd: 0.32 ± 0.06 mg kg^{-1}

P: 116 ± 26 mg kg^{-1}

(coverage factor of 2 for approx. 95% confidence level)

7 Comments

7.1 Contribution of effects

Table A6.6 shows that the sampling/sample preparation process contributes considerably to the overall measurement uncertainty. To recognise and to assess the relevance of single effects/process steps several aspects must be considered:

7.1.1 The *'between-location' effect* depends on the homogeneity of the target area and the total number of increments taken from each square. Former investigations show that 20 increments per hectare of arable land yield an uncertainty contribution in the order of the analytical uncertainty.

7.1.2 The *error due to the sampling strategy* is difficult to quantify, but can often be much larger than that observed in this case study. Practically it can only be controlled by 'expert judgement' of the large-scale distribution of the analyte over the area and the choice of an appropriate sampling strategy.

7.1.3 With the model calculation of the *depth effect*, it is treated as an *unknown* systematic error, that is, the deviation in depth occurs with all increments (more or less) in the same direction. This may be realistic under specific conditions; for example, a dry sandy soil tends to drop out at the lower end of the auger so that the average increment depth would be too small. If such an effect is detected, then the correction of the systematic deviation is possible and only the random error component must be considered (i.e. the uncertainty decreases with the factor of $1/\sqrt{n_{incr}}$). Training of the sampler may reduce this 'point materialisation error'.

7.1.4 The *splitting effect* is hard to control because initial mass reduction is often performed in the field. It can contribute significantly if the method of mass reduction

is inappropriate or performed carelessly. Consequently, training of the sampling personnel is of great importance.

7.1.5 The *effect of moisture content* for air dried soil samples seems to be negligible in this case.

7.1.6 The uncertainty of the *analytical process* can contribute the dominating proportion to the combined measurement uncertainty (e.g. for cadmium). It can be controlled if the standard methods of analytical quality assurance are adhered to (e.g. periodical use of CRMs and participation on inter-laboratory comparisons). Uncertainty from this source may be dominant when the analyte concentration is close to the analytical detection limit.

7.1.7 Effects that were not considered in this case study include the duration and extent of the forces within the grinding and sieving process, and the wetness of the soil body during the sampling process. The influence of these effects was considered not to be significant, although these assumptions should be verified.

8 Assessment of fitness for purpose of these measurements

For a routine measurement according to the sampling protocol one composite sample from approximately 10 increments must be analysed in duplicate.

In this case study for estimation of uncertainty contributions from single effects, 10 additional increments are taken and 20 (composite) samples are prepared and analysed in total.

This additional effort and cost is not appropriate for routine measurements. However, if measurements on arable land are the main type of investigation conducted by the laboratory, such an exploratory investigation might be valuable for getting a typical value of the 'sampling error' component for these measurements. Furthermore, an evaluation of the error components (i.e. uncertainty budget) will also be useful to optimise the measurement process.

9 Reporting and interpretation

Measurements of the mean concentration for this area of top soil have expanded uncertainty values that can be expressed as either 0.06 mg kg^{-1} or 18.2% of the concentration value for cadmium, and 26 mg kg^{-1} or 22.6% for phosphorus.

10 Summary

	Measurement uncertainty*		
Analyte	Sampling	Analytical	Total
Cd	15.0%	10.4%	18.2%
P	11.6%	19.4%	22.6%

* with coverage factor of 2 (i.e. for 95% confidence)

Appendix B: Terminology

Accuracy The closeness of agreement between a test result and the accepted reference value.

Note: The term accuracy, when applied to a set of test results, involves a combination of random components and a common systematic error or bias component.

ISO 3534-1: 3.11 (1993) [9]

Bias The difference between the expectation of the test result and an accepted reference value.

Note: Bias is a measure of the total systematic error as contrasted to random error. There may be one or more systematic error components contributing to the bias. A larger systematic difference from the accepted reference value is reflected by a larger bias value.

ISO 3534-1: 3.13 (1993) [9]

Composite sample (also average and aggregate) Two or more increments/sub-samples mixed together in appropriate portions, either discretely or continuously (blended composite sample), from which the average value of a desired characteristic may be obtained.

ISO 11074-2: 3.10 (1998) [45], AMC (2005) [50]

Duplicate (replicate) sample One of the two (or more*) samples or sub-samples obtained separately at the same time by the same sampling procedure or sub-sampling procedure. *for replicate sample

Note: Each duplicate sample is obtained from a separate 'sampling point' within the 'sampling location'.

Adapted from ISO 11074-2: 2.14 (1998) [45], ISO 1998 was formally adapted from ISO 3534-1 (1993) [9], AMC (2005) [50]

Error of result The test result minus the accepted reference value (of the characteristic).

Note: Error is the sum of random errors and systematic errors.

ISO 3534-1: 3.8 (1993) [9]

Fitness for purpose The degree to which data produced by a measurement process enables a user to make technically and administratively correct decisions for a stated purpose.

Note: As defined for analytical science.

Thompson and Ramsey (1995) [24]

Homogeneity, heterogeneity The degree to which a property or constituent is uniformly distributed throughout a quantity of material.

Notes:

1. A material may be homogeneous with respect to one analyte or property but heterogeneous with respect to another.

2. The degree of heterogeneity (the opposite of homogeneity) is the determining factor of sampling error.

IUPAC (1990) [46]; ISO 11074-2: 1.6 (1998) [45]

Increment Individual portion of material collected by a single operation of a sampling device.

IUPAC (1990) [46], AMC (2005) [50]

Laboratory sample Sample as prepared for sending to the laboratory and intended for inspection or testing.

ISO Standard 78-2 (1999) [47]

Measurand Particular quantity subject to measurement.

ISO-GUM (1993) [2]

Measurement Uncertainty *see* Uncertainty of measurement

Precision

The closeness of agreement between independent test results obtained under stipulated conditions.

Notes:

1. Precision depends only on the distribution of random errors and does not relate to the true value or the specified value.

2. The measure of precision usually is expressed in terms of imprecision and computed as a standard deviation of the test results. Less precision is reflected by a lower standard deviation.

3. 'Independent test results' means results obtained in a manner not influenced by any previous result on the same or similar test object. Quantitative measures of precision depend critically on the stipulated conditions. Repeatability and reproducibility conditions are particular sets of extreme stipulated conditions.

ISO 3534-1: 3.14 (1993) [9]

Primary sample

The collection of one or more increments or units initially taken from a population.

Note: The term primary, in this case, does not refer to the quality of the sample, rather the fact that the sample was taken during the earliest stage of measurement.

IUPAC (1990) [46], AMC (2005) [50]

Random error of result

A component of the error which, in the course of a number of test results for the same characteristic, remains constant or varies in an unpredictable way.

Note: It is not possible to correct for random error.

ISO 3534-1: 3.9 (1993) [9]

Random sample

A sample of *n* sampling units taken from a population in such a way that each of the possible combinations of *n* sampling units has a particular probability of being taken.

ISO 3534-1: 4.8 (1993) [9]

Random sampling; simple random sampling

The taking of *n* items from a lot of *N* items in such a way that all possible combinations of *n* items have the same probability of being chosen.

Notes:

1. Random selection can never be replaced by ordinary haphazard or seemingly purposeless choice; such procedures are generally insufficient to guarantee randomness.

2. The phrase random sampling applies also to sampling from bulk or continuous materials but its meaning requires specific definition for each application.

ISO 7002: A.34 (1986) [48]

Reference sampling

Characterisation of an area, using a single sampling device and a single laboratory, to a detail allowing the set-up of a distribution model in order to predict element concentrations, with known uncertainty, at any sampling point.

IUPAC (2005) [49]

Reference sampling target

The analogue in sampling of a reference material or certified reference material (in chemical analysis).

Note: A sampling target, one or more of whose element concentrations are well characterised in terms of spatial/time variability. The analogue in sampling of a reference material or a certified reference material (in chemical analysis) (notes adapted from IUPAC (2003) draft recommendations; originally defined in ISO Guide 30: 1992).

Thompson and Ramsey (1995) [24]

Representative sample

Sample resulting from a sampling plan that can be expected to reflect adequately the properties of interest in the parent population.

IUPAC (1990) [46], ISO 11074-2: 1.9 (1998) [45], AMC (2005) [50]

Sample

A portion of material selected from a larger quantity of material.

IUPAC (1990) [46], ISO 11074-2 (1998) [45], AMC (2005) [50]

Sample preparation The set of material operations (such as reduction of sizes, mixing, dividing etc.) that may be necessary to transform an aggregated or bulk sample into a laboratory or test sample.

Note: The sample preparation should not, as far as possible, modify the ability of the sample to represent the population from which it was taken.

Adapted from ISO 3534-1: 4.30 (1993) [9]

Sample pre-treatment Collective noun for all procedures used for conditioning a sample to a defined state which allows subsequent examination or analysis or long-term storage.

Adapted from ISO 11074-2: 6.1 (1998) [45]

Sample size Number of items or the quantity of material constituting a sample.

ISO 11074-2: 4.26 (1998) [45], ISO 7002: A.40 (1986) [48]

Sampler Person (or group of persons) carrying out the sampling procedures at the sampling point.

Note: The term 'sampler' does not refer to the instrument used for sampling, i.e. the 'sampling device'.

Adapted from ISO 11074-2 (1998) [45]

Sampling Process of drawing or constituting a sample.

Note: For the purpose of soil investigation 'sampling' also relates to the selection of locations for the purpose of in situ testing carried out in the field without removal of material (from ISO 1998).

ISO 11074-2 (1998) [45], ISO 3534-1 (1993) [9]

Sampling bias The part of the total measurement bias attributable to the sampling.

AMC (2005) [50]

Sampling location The place where sampling occurs within the sampling target. Perhaps used for location within which duplicate (or replicate) samples are taken at particular sampling points.

Sampling plan Predetermined procedure for the selection, withdrawal, preservation, transportation and preparation of the portions to be removed from a population as a sample.

IUPAC (1990) [46], ISO 11074-2 (1998) [45], AMC (2005) [50]

Sampling point The place where sampling occurs within the sampling location. Perhaps used for specific point where duplicate (or replicate) sample taken, within a sampling location.

Note: The accuracy at which a sampling point is located in space or time depends on the surveying method. Duplicate samples are taken from sampling points that reflect this accuracy.

Sampling precision The part of the total measurement precision attributable to the sampling.

AMC (2005) [50]

Sampling procedure Operational requirements and/or instructions relating to the use of a particular sampling plan; i.e. the planned method of selection, withdrawal and preparation of sample(s) from a lot to yield knowledge of the characteristic(s) of the lot.

ISO 3534-1: 4.5 (1993) [], ISO 11704-2 [45] (in part), adopted by AMC (2005) [50]

Sampling target Portion of material, at a particular time, that the sample is intended to represent.

Notes:

1. The sampling target should be defined prior to designing the sampling plan.

2. The sampling target may be defined by Regulations (e.g. lot size).

3. If the properties and characteristics (e.g. chemical composition) of the certain area or period are of interest and must be known then it can be considered a sampling target.

AMC (2005) [50]

Sampling uncertainty *see* Uncertainty from sampling

Sub-sample A sample taken from a sample of a population.

Notes:

1. It may be selected by the same method as was used in selecting the original sample, but need not be so.

2. In sampling from bulk materials, sub-samples are often prepared by sample division. The sub-sample thus obtained is also called a 'divided sample'.

ISO 3534-1: 4.8 (1993) [9]

Sub-sampling (sample division)	Process of selection of one or more sub-samples from a sample of a population. ISO 11074-2 (1998) [45]
Systematic error of result	A component of the error which, in the course of a number of test results for the same characteristic, remains constant or varies in a predictable way. *Note: Systematic errors and their causes may be known or unknown.* ISO 3534-1: 3.10 (1993) [9]
Systematic sampling	Sampling by some systematic method. ISO 3534-1: 4.15 (1993) [9], ISO 11074-2 [45]
Test portion	Quantity of material, of proper size for measurement of the concentration or other property of interest, removed from the test sample. IUPAC (1990) [46], ISO 11074-2: 3.17 (1998) [45], AMC (2005) [50]
Test sample	Sample, prepared from the laboratory sample, from which the test portions are removed for testing or analysis. IUPAC (1990) [46], ISO 11074-2: 3.16 (1998) [45], AMC (2005) [50]
Trueness	The closeness of agreement between the average value obtained from a large series of test results and an accepted reference value. *Notes:* *1. The measure of trueness is usually expressed in terms of bias.* *2. The trueness has been referred to as 'accuracy of the mean'. This usage is not recommended.* ISO 3534-1: 3.12 (1993) [9]

Uncertainty (of measurement)

Parameter, associated with the result of a measurement, that characterises the dispersion of the values that could reasonably be attributed to the measurand.

Notes:

1. The parameter may be, for example, a standard deviation (or a given multiple of it), or the half width of an interval having a stated level of confidence.

2. Uncertainty of measurement comprises, in general, many components. Some of these components may be evaluated from the statistical distribution of the results of series of measurements and can be characterised by experimental standard deviations. The other components, which can also be characterised by standard deviations, are evaluated from assumed probability distributions based on experience or other information.

3. It is understood that the result of the measurement is the best estimate of the value of the measurand, and that all components of uncertainty, including those arising from systematic effects, such as components associated with corrections and reference standards, contribute dispersion.

4. (added) If measurand is defined in terms of the quantity within the sampling target, then uncertainty from sampling is included within uncertainty of measurement.

ISO GUM: B.2.18 (1993) [2]

Uncertainty from sampling

The part of the total measurement uncertainty attributable to sampling.

Note. Also called sampling uncertainty

IUPAC (2005) [49]

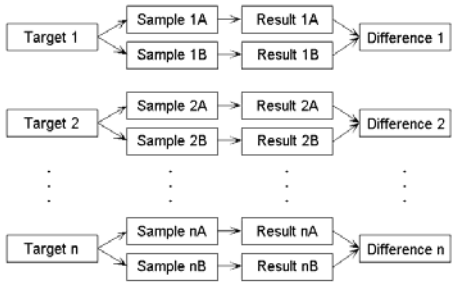
Appendix C: Useful statistical procedures

C1. Estimating bias between two sampling methods, by using paired samples

The paired-sample method is effected by collecting one sample, according to both of the sampling protocols under consideration, from each of a number ($n > 20$ preferably) of targets. The method is especially suitable for comparing a new candidate protocol against an established protocol in routine use, but is also generally applicable. For each method the sampling procedure has to be randomised in some fashion, for example by starting the collection of increments at a random position within the target and orientating the increment grid in a random direction. The samples collected are analysed under randomised repeatability conditions, so that analytical bias is cancelled out.

The design, shown below in Figure C1.1, ensures a minimum of extra work at each target, so that the experiment can be executed at low cost without interrupting the flow of routine sampling. The result is also rugged, because it is derived from data collected from many typical but different targets. It therefore represents the average bias between the results of the two protocols, rather than the bias encountered in a single target, which may turn out to be atypical.

Figure C1.1: Design of experiment to estimate the bias between two sampling methods

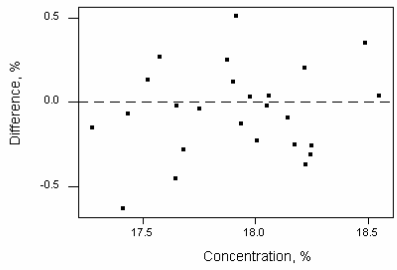


Design of experiment to estimate the bias between two sampling methods A and B, by collecting paired samples at each target.

The first stage of the examination of the results is to check whether the paired differences are dependent on the concentration of the analyte. This is particularly likely to happen if the concentration range encountered in successive targets is wide. A scatterplot provides a useful visual check. Where there is no dependence, the bias estimate is the mean of the signed paired differences and this mean can be tested for significant difference from zero in the usual fashion. In the example shown in Figure C1.2, there is no apparently significant dependence between the signed difference and the concentration, and the bias between the methods is not significantly different from zero at the 95% level of confidence by the two-sample *t*-test. Where there is a clear bias that is dependent on concentration, as in Figure C1.3, the bias should be expressed as a function of concentration. In the instance illustrated, there is evidence (established by the functional relationship method [40]) of a significant rotational bias with a trend expressed by the equation Result (B) = Result (A) × 1.2.

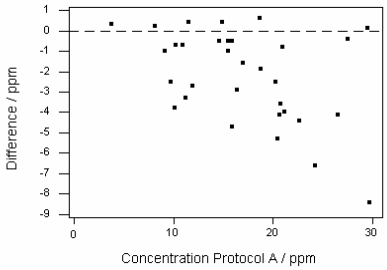
.

Figure C1.2: No significant bias or trend



Differences between results of two sampling protocols applied to 25 targets, as a function of the concentration. There is no significant bias and no suggestion of a dependence of bias on concentration.

Figure C1.3: Significant bias and trend



Differences between results of two sampling protocols applied to 33 targets, plotted as a function of the concentration. There is a significant bias (because 27/33 results are negative) and the absolute bias increases with increasing concentration.

C2. Further description of sampling errors from sampling theory

C2.1 Weighting error (SWE) forms its own class. It is created, for example, if the lot (sampling target) consists of sub-lots of different sizes but the mean concentration is estimated as a simple mean, without taking the sizes of the sub-lots into account. The correct method is to calculate the weighted mean by using the sizes of the sub-lots as weights. In analysis of moving material, weighting error is generated if the flow-rate varies but is not taken into account in calculating the mean; in this case the flow-rates should be recorded simultaneously with sampling and used as weights in calculating the mean. Another option is to use a sampling device that cuts samples whose size is proportional to the flow-rate and use the sample sizes as weights in calculating the mean. It should be noted that if a composite sample is made from sub-samples then proportional sampling should be used; otherwise a weighting error is generated in the composite sample.

C2.2 Grouping and segregation error (GSE) is the second error term related to short range errors. It is caused by the fact that the sample is normally not taken fragment by fragment, but as a group of fragments. If there is segregation in the material, this causes this type of error. This error is not normally estimated. Gy has shown, however, that if the sampling is correctly done GSE is smaller than, or at maximum equal to, the fundamental sampling error (FSE).

C2.3 Point selection error (PSE). When the mean of a continuous object (e.g. process stream, river, polluted site, ...) is estimated by using discrete samples, the uncertainty of the mean depends on the sampling strategy, because the results are usually *autocorrelated*. This error is called point selection error (PSE) and it depends on the sampling strategy. Three basic strategies can be applied for picking the samples (see Figure C2.1):

1) **Random sampling:** Time or location of the sampling points are randomly distributed along the target.
2) **Stratified (random) sampling:** The lot is first divided into N sub-lots of equal sizes and within each sub-lot the sampling point is randomly assigned.
3) **Systematic (stratified) sampling:** All N samples are collected at equal distances (one-dimensional case) or on a fixed symmetric pattern (targets which from the sampling point of view have two or more dimensions).

Estimation of the standard deviation of the mean of the lot

Random sampling**:** $s(a_L) = \frac{s_p}{\sqrt{N}}$

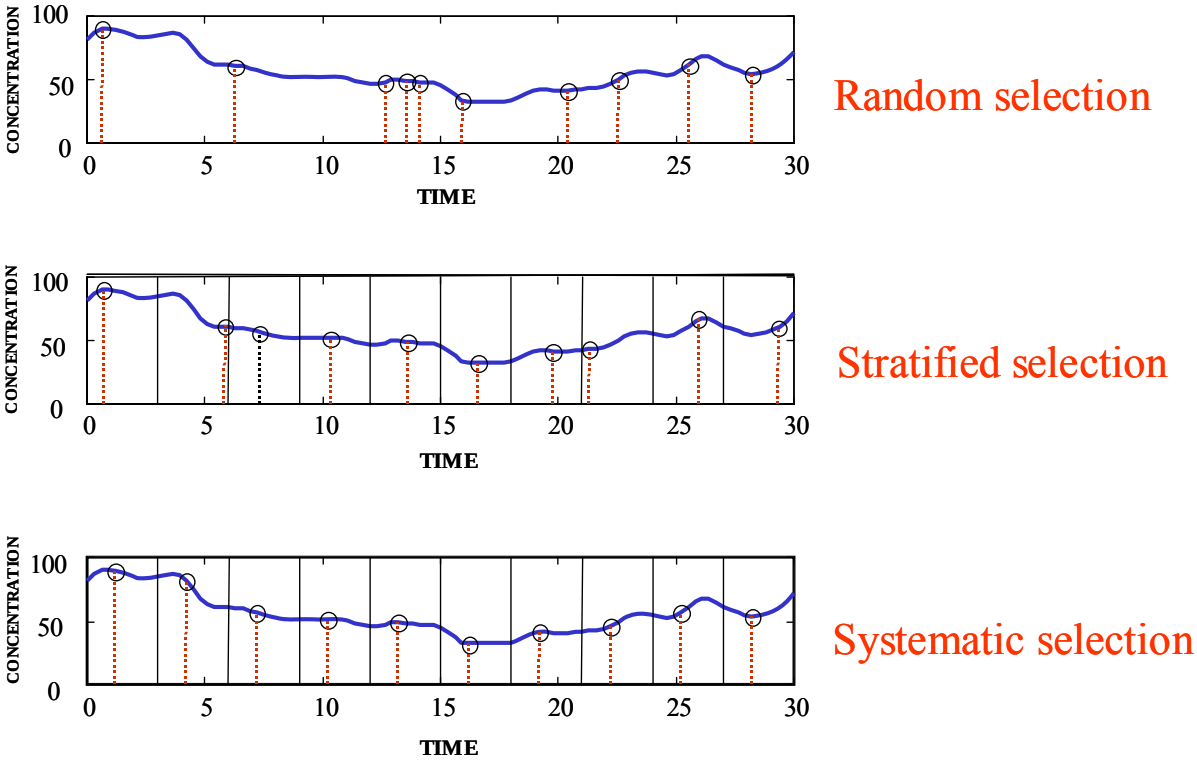
Stratified sampling**:** $s(a_L) = \frac{s_{strat}}{\sqrt{N}}$

Systematic sampling**:** $s(a_L) = \frac{s_{sys}}{\sqrt{N}}$

s_{strat} and s_{sys} are standard deviation estimates, where the autocorrelation has been taken into account.

Normally the order is $s_p > s_{strat} > s_{sys}$, except when in systematic sampling the sampling frequency is a multiple of process frequency. In this case the systematic sampling is the worst choice and the mean may be biased.

Figure C2.1: Sampling strategies



Ten samples selected from the target by using *random*, *stratified* random and *systematic* stratified sample selection.

Estimation of the PSE

The distribution heterogeneity of a one-dimensional lot can be characterised by carrying out a variographic experiment, i.e. N samples are collected from the target by using systematic sample selection. N should be at least 30, preferably 60…100. Proportional cross-stream sampling should be used or if not possible (when large gas or liquid streams are sampled) the flow-rate should be recorded simultaneously with the sampling time. From these results the experimental heterogeneity h_i can be calculated as the relative variation about the lot mean (or mean of the sampling target). When N samples of size M_i are collected and analysed (results are a_i). M_i can be also the flow-rate, if proportional sampling cannot be carried out.

$$h_i = \frac{a_i - a_L}{a_L}\frac{M_i}{\overline{M}} \quad (i = 1, 2, \ldots, N)$$

where a_L is the weighted mean of the lot:

$$a_L = \frac{\sum M_i a_i}{\sum M_i} = \frac{1}{N}\sum\left(\frac{M_i}{\overline{M}}\right)a_i$$

The standard deviation of the heterogeneity h is equal to the *relative standard deviation* of the lot or process, s_p.

To characterise the variability of the process an experimental variogram is calculated from the heterogeneities:

$$V_j = \frac{1}{2(N-j)} \sum_{i=1}^{N-j} \left(h_{i+j} - h_i\right)^2, \qquad j = 1,2,\ldots,\frac{N}{2}$$

The variogram has to be integrated to estimate the PSE for different sampling strategies. Gy uses a robust numerical integration.

C3. Sources of software for calculations

Classical analysis of variance (ANOVA) is available in most general spreadsheet software for one-way ANOVA. F-tests and other standard statistical tests for the normal distribution are also implemented in most spreadsheets.

Programs for general robust statistical methods in general, and for robust ANOVA in particular, are available from RSC/AMC (http://www.rsc.org/Membership/Networking/InterestGroups/Analytical/AMC/Software/index.asp).

Outlier tests (e.g. Grubb's or Dixon's) are less generally available, as is software for the range method. The range method can, however, be implemented relatively simply using maximum and minimum functions in a spreadsheet.

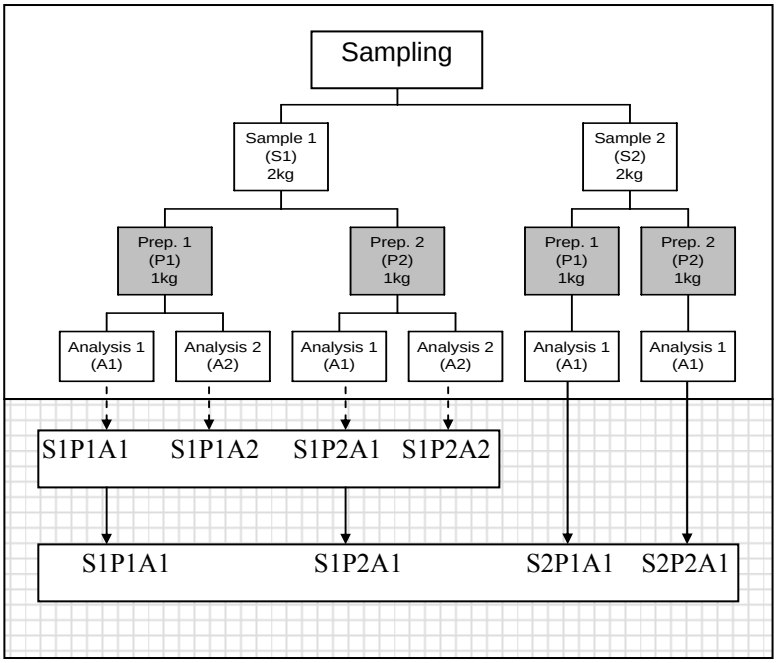
The range calculations (demonstrated in Section 7 of Appendix A3) are easily performed using standard spreadsheets, and an example can be downloaded from http://team.sp.se/analyskvalitet/sampling/default.aspx.

Appendix D: Alternative experimental designs for empirical uncertainty estimation

1. Multi-level designs to estimate other component effects

The general balanced design for the empirical estimation of uncertainty (Figure 2) includes the uncertainty from the physical sample preparation in with the 'sample' step. An alternative experimental design (Figure D.1) can be used to make a separate estimate of the uncertainty from this source (s_{prep}). Two sub-samples from both of the two primary samples are prepared separately (grey boxes in Figure D.1). Duplicate test portions are taken from these sub-samples so that the analytical contribution can also be estimated. The standard robust ANOVA can be used to separate all of these sources of variance (Figure A1.2, and Appendix C3), by selecting two different sub-sets of four measurements, shown in Figure D.1. Full details of the application of this design to food sampling are given elsewhere [30].

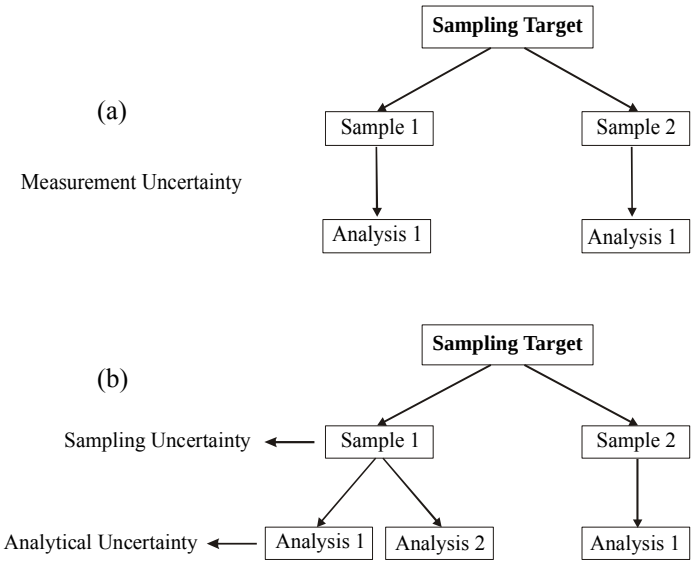
Figure D.1: Experimental design utilised for the estimation of uncertainty from sample preparation, as well as that from sampling and analysis



* The upper section depicts the three-layered and unbalanced experimental design. The additional layer in this experimental design, required for the evaluation of s_{prep}, is shown by the grey boxes. The lower section (shaded) shows the data groupings required for the application of ANOVA so as to provide estimates of s_{samp}, s_{prep} and s_{anal}, i.e. the statistical design. Figure taken from [30] with permission of Royal Society of Chemistry.

2. Simplified and unbalanced designs, to reduce the cost of implementation

Figure D.2 Two simplified alternatives to the full balance design (Figure 2) that can be applied to reduce the cost of estimating the measurement uncertainty using the empirical approach: (a) the simplified balanced design, and (b) the unbalanced design



The simplified design (Figure D.2a) has the same duplicated samples as in the full balanced design (Figure 2), but does not include and duplicated chemical analyses. The uncertainty estimated using this design gives the total measurement uncertainty, without any values for the components of the uncertainty from the sampling or the analysis. If these components are required, the analytical uncertainty can be estimated externally by the laboratory, and removed from the total uncertainty, to give a separate estimate of the sampling uncertainty, using Equation 1. The main advantage of this design is that the analytical cost of implementation is only half of that for the full balanced design, for the same number of duplicated samples. Alternatively, twice the number of duplicated samples can be taken, from twice the number of targets to increase their representativeness for the same expenditure on chemical analysis.

The unbalanced design (Figure D.2b) is intermediate between these two designs, with only one analytical duplicate carried out on one of the duplicated samples. This has the advantage of giving estimates of the sampling and analytical components of the uncertainty, as well as the total measurement uncertainty (with the same caveats expressed as for the full balanced design in Section 9.4.2). The analytical costs are reduced by 25% compared with those for the fully balanced case. The degrees of freedom in this case are similar for both the analytical and sampling estimates of variance, which is more cost-effective than the extra degrees of freedom for the analytical uncertainty in the fully balanced case.

Classical ANOVA can be applied to the output of both of these designs using many different spreadsheet software packages (Appendix C3), but robust ANOVA has not yet been developed for this case.

Appendix E: Modifying sampling uncertainty using predictions from sampling theory

Once uncertainty from sampling has been estimated, and if it is found not to be fit for purpose, there may be a need to modify this level of uncertainty. Predictions on how to achieve this modification can be made using sampling theory (Section 10.2). Several theories predict that the sampling variance is inversely proportional to the mass of the sample taken (e.g. Equation 5). This leads to the prediction that any required modification of uncertainty of sampling (from u_{samp1} to u_{samp2}) can be calculated by changing the mass of the sample (from m_{s1} to m_{s2}) using the relationship

$$m_{s2} = (u_{samp1} / u_{samp2})^2 \, . \, m_{s1} \qquad \text{……..(Equation E1)}$$

This approach can usefully be illustrated using the case study of nitrate in lettuce in Example A1. The sampling uncertainty was shown not to be fit for purpose (by the method in Section 16.3), and the optimal uncertainty required was calculated to be lower by a factor of approximately 2. Equation E1 predicts that this should be achieved by increasing the sample mass by a factor of 4 (i.e. 2^2). The implementation of this prediction by increasing the number of increments from 10 heads to 40 heads of lettuce per batch, did achieve the predicted reduction in the sampling uncertainty in this case (i.e. by a factor of 1.80, which is not statistically significantly different from the predicted improvement of 2.0) [38]. Such successful predictions are not always achieved in practice. In a different example for the determination of moisture in butter, a predicted reduction of 3.7 in the u_{samp}, was calculated to require an increase in m_s by a factor of 14. In practice this increase in sample mass only produced an experimental improvement of 1.3. The inability of this model to predict the change in sampling uncertainty was probably due to the nature of the heterogeneity of the analytes in this particular material [51].

Appendix F: References

1 Ellison S L R, Roesslein M, Williams A (eds) (2000) Eurachem/CITAC Guide: *Quantifying Uncertainty in Analytical Measurement*, Eurachem, 2nd edition, ISBN 0 948926 15 5. Available from the Eurachem secretariat, or from LGC Limited (London).

2 ISO (1993) *Guide to the Expression of Uncertainty in Measurement* (GUM). Geneva (2nd printing 1995).

3 Ellison S L R, Roesslein M, Williams A (eds) (1995) Eurachem Guide: *Quantifying Uncertainty in Analytical Measurement*, Eurachem, 1st edition.

4 Analytical Methods Committee (1995) Uncertainty of measurement: implications of its use in analytical science. *Analyst*, 120, 2303–2308.

5 ISO/TS 21748 (2004) Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty estimation. ISO, Geneva.

6 Gy P M (1979) *Sampling of Particulate Materials – Theory and Practice*. Elsevier, Amsterdam, 431pp.

7 Codex (2004) General Guidelines on Sampling. CAC/GL-2004 (http://www.codexalimentarius.net/download/standards/10141/CXG_050e.pdf)

8 Nordtest (2007) Uncertainty from sampling. A Nordtest handbook for sampling planners and sampling quality assurance and uncertainty estimation. NT tec 604/TR604 (www.nordicinnovation.net)

9 ISO 3534-1: 1993 Statistics – Vocabulary and Symbols, International Organization for Standardization, Geneva

10 Ramsey M H, Squire S, Gardner M J (1999) Synthetic reference sampling target for the estimation of measurement uncertainty. *Analyst*, 124 (11), 1701–1706.

11 Ramsey M H (1998) Sampling as a source of measurement uncertainty: techniques for quantification and comparison with analytical sources. *Journal of Analytical Atomic Spectrometry*, 13, 97–104.

12 Lyn J A, Ramsey M H, Coad D S, Damannt A P, Wood R, Boon K A The duplicate method of uncertainty estimation: are 8 targets enough? *(Submitted to Analyst)*

13 De Zorzi P, Belli M, Barbizzi S, Menegon S, Deluisa A (2002) A practical approach to assessment of sampling uncertainty. *Accreditation and Quality Assurance*, 7, 182–188.

14 Kurfurst U, Desaules A, Rehnert A, Muntau H (2004) Estimation of measurement uncertainty by the budget approach for heavy metal content in soils under different land use. *Accreditation and Quality Assurance*, 9, 64–75.

15 Minkkinen P (2004) Practical applications of sampling theory. *Chemometrics and Intelligent Lab. Systems*, 74, 85–94.

16 Gy P M (1998) *Sampling for Analytical Purposes*. John Wiley & Sons Ltd, Chichester.

17 Gy P M (1992) *Sampling of Heterogeneous and Dynamic Material Systems*. Elsevier, Amsterdam.

18 Gy P M (2004) Proceedings of First World Conference on Sampling and Blending. Special Issue of *Chemometrics and Intelligent Lab. Systems,* 74, 7–70.

19 Pitard F F (1993) *Pierre Gy's Sampling Theory and Sampling Practice.* CRC Press, Boca Raton, 2nd edition.

20 Smith P L (2001) *A Primer for Sampling Solids, Liquids and Gases – Based on the Seven Sampling Errors of Pierre Gy*. ASA SIAM, USA.

21 Ramsey M H (1993) Sampling and analytical quality control (SAX) for improved error estimation in the measurement of heavy metals in the environment, using robust analysis of variance. *Applied Geochemistry*, 2, 149–153.

22 Nordtest Sampler Certification, Version 1-0, Nordic Innovation Centre (2005).

23 Thompson M, Wood R (1995) Harmonised guidelines for internal quality control in analytical chemistry laboratories. *Pure and Applied Chemistry*, 67, 649–666.

References (Continued)

24 Thompson M, Ramsey M H (1995) Quality concepts and practices applied to sampling – an exploratory study. *Analyst*, 120, 261–270.

25 Magnusson B, Naykki T, Hovind H, Krysell M (2003) *Handbook for Calculation of Measurement Uncertainty*. NORDTEST report TR 537 (project 1589-02). Can be obtained from www.nordicinnovation.net/nordtest.cfm under link Rapporter.

26 Ramsey M H, Squire S, Gardner M J (1999) Synthetic reference sampling target for the estimation of measurement uncertainty. *Analyst*, 124 (11), 1701–1706.

27 Thompson M, Coles B J, Douglas J K (2002) Quality control of sampling: proof of concept. *Analyst*, 127, 174–177.

28 Farrington D, Jervis A, Shelley S, Damant A, Wood R, Thompson M (2004) A pilot study of routine quality control of sampling by the SAD method, applied to packaged and bulk foods. *Analyst*, 129, 359–363.

29 Codex (2004) Guidelines on measurement uncertainty CAC/GL 54-2004.

30 Lyn J A, Ramsey M H, Fussel RJ, Wood R (2003) Measurement uncertainty from physical sample preparation: estimation including systematic error. *Analyst*, 128 (11), 1391 – 1398

31 *Uncertainty and Assessment of Compliance* Eurachem (in preparation).

32 *Measurement Uncertainty and Conformance Testing: Risk Analysis.* Joint Committee on General Metrology, Working group 1/Subcommittee 3 (in preparation).

33 Ramsey M H, Thompson M, Hale M (1992) Objective evaluation of precision requirements for geochemical analysis using robust analysis of variance. *Journal of Geochemical Exploration*, 44, 23–36.

34 Thompson M, Fearn T (1996) What exactly is fitness for purpose in analytical measurement? *Analyst*, 121, 275–278.

35 Commission Regulation (EC) No 563/2002 of 2 April 2002 amending Regulation (EC) No 466/2001 setting maximum levels for certain contaminants in foodstuffs, Official Journal of the European Communities, L 86/5 to L 86/6.

36 European Directive 79/700/EEC. OJ L 207, 15.8.1979, p26.

37 Ramsey M H, Lyn J A, Wood R (2001) Optimised uncertainty at minimum overall cost to achieve fitness-for-purpose in food analysis. *Analyst*, 126, 1777–1783.

38 Lyn, J A, Palestra I M, Ramsey M H, Damant A P, Wood R (2007) Modifying uncertainty from sampling to achieve fitness for purpose: a case study on nitrate in lettuce. *Accreditation and Quality Assurance: Journal for Quality, Comparability and Reliability in Chemical Measurement,* 12*,* 67-74.

39 Thompson M, Walsh J N A (1989) *Handbook of Inductively Coupled Plasma Spectrometry*. Blackie, Glasgow.

40 Analytical Methods Committee (2002) Technical Brief No.10, Royal Society of Chemistry, London.

41 Ramsey M H, Taylor P D, Lee J C (2002) Optimized contaminated land investigation at minimum overall cost to achieve fitness-for-purpose. *Journal of Environmental Monitoring*, 4 (5), 809–814.

42 Lee J C, Ramsey M H (2001) Modelling measurement uncertainty as a function of concentration: an example from a contaminated land investigation. *Analyst*, 126 (10), 1784–1791.

43 Ramsey M H, Argyraki A (1997) Estimation of measurement uncertainty from field sampling: implications for the classification of contaminated land. *Science of the Total Environment*, 198, 243–257.

44 Dahinden R, Desaules A (1994) *Die Vergleichbarkeit von Schwermetallanalysen in Bodenproben von Dauerbeobachtungsflächen, Eidgenössische Forschungsanstalt für Agrikulturchemie und Umwelthygiene*. Liebefeld, Switzerland, p27.

45 ISO 11074-2: 1998 Soil Quality – Vocabulary. Part 2: Terms and definitions related to sampling, International Organization for Standardization, Geneva.

46 IUPAC (1990) Nomenclature for sampling in analytical chemistry (Recommendations 1990), prepared for publication by Horwitz W, *Pure and Applied Chemistry*, 62, 1193–1208.

References (Continued)

47 ISO Standard 78-2: Chemistry – Layouts for Standards – Part 2: Methods of Chemical Analysis (Second Edition, 1999).

48 ISO 7002: 1986 Agricultural food products – Layout for a standard method of sampling from a lot, First Edition, International Organization for Standardization, Geneva.

49 IUPAC (2005) Terminology in Soil Sampling (IUPAC Recommendations 2005), prepared for publication by De Zorzi P, Barbizzi S, Belli M, Ciceri G, Fajgelj A, Moore D, Sansone U, and Van der Perk M. *Pure and Applied Chemistry*, 77 (5), 827–841.

50 AMC (2005) Analytical Methods Committee Technical Brief No 19. Terminology – the key to understanding analytical science. Part 2: Sampling and sample preparation (http://www.rsc.org/images/brief19_tcm18-25963.pdf).

51 Lyn J A, Ramsey M H, Damant A, Wood R (2005), Two stage application of the OU method: a practical assessment, *Analyst,* 130, 1271-1279

ISBN 978 0 948926 26 6