Eurachem

CITAC
Cooperation on International Traceability in Analytical Chemistry

EURACHEM / CITAC Guide

Validation of Measurement Procedures that Include Sampling

Supplement to:
"The Fitness for Purpose of Analytical Methods"
"Measurement Uncertainty arising from Sampling"

First Edition 2024

Produced jointly with:
Eurachem, EUROLAB, CITAC, Nordtest
RSC Analytical Methods Committee

Validation of Measurement Procedures that Include Sampling (VaMPIS)

Supplement to Eurachem Guides
The Fitness for Purpose of Analytical Methods and
Measurement Uncertainty arising from Sampling

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

Editors

Michael H. Ramsey (University of Sussex, UK)
Peter D. Rostron (University of Sussex, UK)
Fernando C. Raposo (Portugal)

Composition of the Joint Task Group*

Eurachem members

Michael H. Ramsey (Co-Chair): University of Sussex, UK
Marios Kostakis (Co-Chair): National and Kapodistrian University of Athens, Greece
Fernando C. Raposo: Portugal
Patrizia Stefanelli: Istituto Superiore di Sanità, National Institute of Health, Italy
Vassilis Zonaras: Directorate of Pesticides' Control and Phytopharmacy, Kifissia, Attica,
Alex Williams: Eurachem, UK
Perihan Yolci Ömeroğlu: Bursa Uludag University, Food Engineering Department, Turkey
Nathalie Guigues: Laboratoire national de métrologie et d'essais, France

EUROLAB members

Andrea Paul: Bundesanstalt für Materialforschung und – prüfung (BAM), Germany
Fulvio Mattaliano: Chemical Laboratory of TUV SUD, Italy
Tatjana Tomić: Croatian Measurement Society, Croatia
Sandra Babić: University of Zagreb, Faculty of Chemical Engineering and Technology, Croatia

CITAC member

Ricardo Bettencourt de Silva: University of Lisboa, Portugal

Nordest member

Bertil Magnusson: Trollboken AB, Sweden

RSC AMC representative

Peter Rostron (Secretary): University of Sussex, UK
*At time of document approval

Acknowledgements

This document has been produced primarily by a joint Eurachem/EUROLAB/ AMC Working Group with the composition shown (right) working in the period 2021–2024. The editors are grateful to all these individuals and also to organisations and others who have contributed comments, advice and assistance.

Lorens Sibbesen, Labquality International (DK) (up to August 2023).

Particular thanks to Laura Martin and Martina Razzaboni (EUROLAB) for organising and hosting an online discussion forum.

This publication should be cited as:
M. H. Ramsey, P. D. Rostron, F. C. Raposo (eds.)
Eurachem/EUROLAB/CITAC/Nordtest/AMC Guide: *Validation of Measurement Procedures that Include Sampling, Eurachem* (2024). Available from https://www.eurachem.org

Foreword

The recognition that the measurement process starts when a **primary sample** is taken from the material the **sample** is intended to represent (i.e. the **sampling target**), requires that the **validation** process should also begin at that point. Traditionally, primary **sampling** and chemical (or physical) analysis were regarded as two separate activities, often conducted by different organisations with staff largely unaware of the activities of the other, and with different approaches to ensuring data quality. The validation of analytical methods, or analytical procedures, when considered in isolation, is well established [1]. However, the new appreciation of the integrated nature of the overall measurement process, gives an opportunity to bring these two activities of sampling and analysis together in a unified approach to validation. Many of the performance characteristics used in analytical method validation, e.g. **working range**, **analytical sensitivity**, **selectivity**, are not generally applicable to **sampling procedures**, but the uncertainty associated with the resultant measurement value, although not currently recognised as a characteristic of the **measurement procedure**, is actually the most important for judging the **fitness for purpose** and hence the validity of the overall measurement procedure.

This new guidance describes an extension to the established validation methodologies to give an integrated approach for the overall measurement process, and hence the *Validation of Measurement Procedures that Include Sampling* (VaMPIS). It does so by making a more realistic estimate of the **measurement uncertainty** [2], including the measurement uncertainty that arises from sampling [3], and uses that as a key metric to judge the fitness for purpose of the measurement procedure and hence the resultant measurement values. This approach is equally applicable to measurements made *in situ* (i.e. without extraction of a sample) and those made *ex situ* (i.e. in a testing laboratory, upon delivery of a previously extracted sample). Two worked examples are used to explain how this approach can be applied for both *in situ* and *ex situ* measurement procedures.

Development of this guidance therefore required collaboration between specialists in the two areas of analytical method validation and measurement uncertainty, especially the measurement uncertainty component arising from sampling (i.e. sampling uncertainty). It also benefited from communication with staff working for organisations that carry out both sampling and/or analysis, for example via an online discussion forum (organised by EUROLAB).

Summary

This Guide aims to explain how to validate an overall measurement procedure from the moment that a primary sample is selected (and usually extracted) from a particular sampling target, until the reporting of a measurement result. The reliable interpretation of the measurement result, such as for assessment of **compliance against limits** [2], requires not just a measurement value but also a realistic estimate of the measurement uncertainty associated with it. Ultimately, it is the measurement uncertainty that summarises the quality of the measurement value, and brings together the contributions from all components of the measurement procedure (sampling and analysis). The measurement uncertainty can therefore be used to validate the overall measurement procedure by judging whether it is fit for its particular intended purpose (Section 1), provided that all other performance characteristics of the analytical portion of the measurement procedure are also demonstrated to be fit for that purpose.

An integrated approach is taken to the validation in which the sampling, **sample preparation** and analytical steps, are all considered as component parts of the overall measurement process. Where a target measurement uncertainty is already specified for a particular combination of analyte and sampling target, the actual overall measurement uncertainty, and its main components, can be estimated by approaches such as the *Duplicate Method* together with estimation of analytical **bias** using certified reference materials (CRMs) [3]. The overall measurement uncertainty can then be compared against a *target uncertainty* to judge the fitness for purpose of the overall measurement procedure. Options for setting a target uncertainty for a particular situation are described. If the actual measurement uncertainty is unacceptably above (or below) the target uncertainty, then the components of the measurement uncertainty (e.g. sampling, sample preparation and analysis) can be reviewed (as well as their relative costs) to identify which component(s) can best be modified to reach the target uncertainty and hence achieve fitness for purpose (Section 2).

Two worked examples are given to explain how this approach to VaMPIS can be applied. In the first of these examples (measurement of the nitrate concentration in field lettuce) a *sequential* approach is applied. A previously validated *ex situ* analytical procedure (i.e. analytical method) is used to judge the fitness for purpose of an overall measurement procedure that includes field sampling (according to a European Union recommended procedure), hence 'integrating' a sampling procedure into an existing (currently only analytical) measurement procedure. The target uncertainty for this situation is not specified externally, so it is calculated using the *Optimised Uncertainty*[a] methodology. The existing measurement procedure is shown to be not fit for purpose, mainly due to the sampling component of the measurement uncertainty. It is shown that increasing the number of lettuce heads used in the **composite sample** from 10 to 40 **increments** can reduce the measurement uncertainty to a value that is much closer to the target value. Additional reductions in the analytical component of the measurement uncertainty, below the level considered to be fit for purpose when considered in isolation, can also further help achieve fitness for purpose of the overall measurement procedure.

The second example takes a *simultaneous (integrated)* approach to validate the overall *in situ* measurement procedure (i.e. sampling and analysis) for lead in soil using a Portable X-ray Fluorescence Spectrometry (pXRF) device. The random component of the measurement uncertainty is estimated using the Duplicate Method, and the systematic component by a comparison between the *in situ* and *ex situ* measurement values on the same sampling targets, in addition to analyses of an appropriated CRM. Two different approaches for setting fitness for purpose criteria, for two different intended purposes, are compared. The *in situ* measurement procedure is found to be fit for the purpose of geochemical mapping, but not for judging compliance against a regulatory threshold, at this site.

[a] Optimised Uncertainty is *a value of measurement uncertainty that minimises the total cost of both making the measurement (including sampling) and the consequences that arise as a result of the effects of that uncertainty on compliance decisions.*

The need for ongoing **quality control** (QC, internal and external) of the routine application of validated measurement procedures is discussed (Section 3). The management of the overall measurement procedure is essential in order to apply VaMPIS effectively. Traditionally, some organisations that undertake primary sampling have operated with little communication with the analytical laboratories who undertake the analytical procedure and report the measurement results. In order to validate the overall measurement procedure, and apply ongoing QC, there needs to be an increased level of communication and effective cooperation between all organisations responsible for these activities (Section 4).

CONTENTS

List of acronyms and abbreviations

ANOVA	analysis of variance
AQC	analytical quality control
BCR	Community Bureau of Reference
CRM	certified reference material
CTS	collaborative trial in sampling
Df	degrees of freedom
EoL	expectation of loss
FPAM	The Fitness for Purpose of Analytical Methods (Eurachem guide)
GPS	Global Positioning System
GUM	Guide to the Expression of Uncertainty in Measurement
HPLC	high performance liquid chromatography
ICP-AES	inductively coupled plasma atomic emission spectroscopy
IMQC	integrated measurement quality control
IMQCP	integrated measurement quality control plan
ISO	International Organization for Standardization
LOD	limit of detection
LOQ	limit of quantification
MPT	measurement proficiency testing
NBS	National Bureau of Standards
NIST	National Institute of Standards and Technology
pXRF	portable x-ray fluorescence (spectrometer)
PT	proficiency testing
QA	quality assurance
QC	quality control
QUAM	Quantifying Uncertainty in Analytical Measurement (Eurachem guide)
RANOVA	robust analysis of variance
RM	reference material
RST	reference sampling target
SPT	sampling proficiency testing
SS	sum of squares
VaMPIS	Validation of Measurement Procedures that Include Sampling

1 Introduction

1.1 Rationale for the Guidance

Validation of a measurement procedure has traditionally focussed on the analytical component of that process, which usually occurs in the laboratory (i.e. *ex situ*). However, there has been an increasing realisation that the measurement process actually begins at the time that the primary sample is selected from the sampling target [3]. This realisation becomes even clearer when an *in situ* device is used to make a measurement at the sampling target, when the previously apparently separate steps of sampling and analysis combine into one measurement procedure. The implications of this realisation for the validation process are that it must be redesigned to include all of the steps of the measurement procedure, including sampling and any physical preparation or preservation applied to that primary sample. Rather than trying to simply add sampling to the traditional approach for analytical validation, it is more effective to apply a fresh *integrated approach* to the validation of the overall measurement procedure. This integrated approach uses the uncertainty of the measurement value as the key metric that unites and quantifies the effects of *all* of the steps in the measurement procedure (sampling and analysis).

1.2 Aim and intended audience of the Guidance

The primary aim of this Guide is to explain the approaches that can be used to validate an *overall* measurement procedure that includes the primary sampling, as well as to highlight the importance of ongoing quality control and management issues. The primary audience is intended to be those who design and validate measurement procedures as an overall process (*in situ* or *ex situ*), and in particular for those who design sampling procedures. It will also be a useful background for those who routinely implement measurement procedures, and monitor ongoing quality assurance on analytical measurements, including both sampling and analytical components.

1.3 Context of Guidance

1.3.1 General context

Primary sampling and analysis (testing) are often undertaken by different organisations. The historical separation between people and organisations involved in these two components of the overall measurement process creates several challenges, discussed below.

1.3.2 For validation

The concept and practice of validation is now well established in testing (e.g. analytical) laboratories but has previously rarely been formally applied to include the sampling procedure. Validation of a sampling procedure in isolation, without integrating it with the subsequent analytical measurement process or testing, is potentially attractive to organisations only involved in sampling, and has been accredited by some accreditation bodies. However, the evidence for fitness for purpose cannot be made quantitative or objective without measurement values, as required by the ISO/IEC 17025 [4] standard and by some accreditation bodies. Where a single organisation undertakes the overall measurement process (i.e., sampling and measurement, whether quantitative or qualitative [5]), then it is relatively straightforward to validate the overall process. However, when two (or more) organisations are involved, then there needs to be ongoing communication and cooperation between them in order to conduct the validation effectively.

1.3.3 For regulators

For regulators, there needs to be an awareness that reliable compliance decisions need to be based on measurement results from validated measurement procedures (including both sampling and analysis/testing). The traditional assumption that a sample can be considered fully ‘representative of a

given sampling target' (hence assuming a negligible sampling uncertainty component of the overall measurement uncertainty), if it is taken 'correctly' by a 'correct' sampling procedure, needs to be critically evaluated.

1.3.4 For accreditation bodies

For accreditation bodies, the definition of the measurement procedure needs to be extended to include the primary sampling and all of the steps that can occur outside of the testing laboratory (e.g., sample preservation, sample transportation and its physical preparation, as required by the ISO/IEC 17025, Chapter 7.4.1 [4]. Quantitative evidence needs to be sought to demonstrate the validity of the overall measurement procedure, hence including sampling and analysis.

1.4 Eurachem guide FPAM: 'The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation'

The Eurachem FPAM Guide [1] is a well-established document that describes the steps required to validate an analytical method (or more correctly the measurement procedure [6]) that must be subjected to validation, by judging the fitness for purpose of the measurement results. It uses estimates for eight performance characteristics[b] of the analytical procedures that occur after the 'laboratory sample' has been delivered to the testing laboratory (Figure 1, final 3-4 steps shaded pale grey or unshaded). The fourth step of physical preparation of the laboratory sample usually occurs in a laboratory but is not always included in the analytical validation process.

1.5 Eurachem Guide: 'Measurement Uncertainty arising from Sampling'

The Eurachem Sampling Uncertainty Guide [3] primarily considers estimation of the measurement uncertainty contribution due to the sampling process (sampling uncertainty) in the context of the overall measurement process (Figure 1, first 3-4 steps, shaded darker grey). The measurement process is considered to start when the primary sample is taken from the sampling target (e.g. a batch, lot or volume of material), and to end when the analytical measurement result is reported. For validation of the overall measurement process, the overall measurement uncertainty that arises from both components, sampling and testing (e.g. chemical analysis), is identified as the key unifying metric that can be used to decide the fitness for purpose of the resultant measurement values, and hence to achieve its quantitative and transparent validation.

1.6 Purpose of starting a measurement process

The purpose of a measurement process (including sampling and testing, such as chemical analysis) is to enable the user of the measurement results to make reliable decisions. For example, a measurement result may be used in a compliance (or conformity[c]) assessment to decide whether the analyte concentration in a sampling target is below (or above) a given regulatory limit. Each measurement result is usually composed of two values, the estimated value of concentration and its associated uncertainty. The measurement uncertainty must be estimated reliably and be small enough to be able to make a reliable decision, but not so small that it makes the measurement procedure disproportionately expensive. A measurement procedure can be considered validated if it can be shown to be fit for purpose. Fitness for purpose is often defined in terms of a target uncertainty. Generally, the target uncertainty

[b] Selectivity, limit of detection (LOD) and limit of quantification (LOQ), working range, analytical sensitivity, trueness (bias, recovery), precision (repeatability, intermediate precision and reproducibility), measurement uncertainty, ruggedness (robustness).

[c] The term 'compliance assessment' is generally used in this document, but the techniques would usually be equally applicable to 'conformity assessment'.

(that should include the sampling uncertainty component) should be set in a dedicated regulation, or agreed upon between a testing laboratory and its customer. One option for setting a target measurement uncertainty, if required, is to use the optimum measurement uncertainty, which minimises the cost of both the measurement and the potential costs of misclassification (e.g. due to an incorrect compliance statement, see Appendix A, examples A1, A2 and Appendix B). The two main components of the measurement procedure, sampling and analytical, both contribute to the overall measurement uncertainty. The validation needs to look primarily at the overall measurement uncertainty, but subsequently also at these two components, enabling the target level of measurement uncertainty to be achieved in the most cost-effective way.

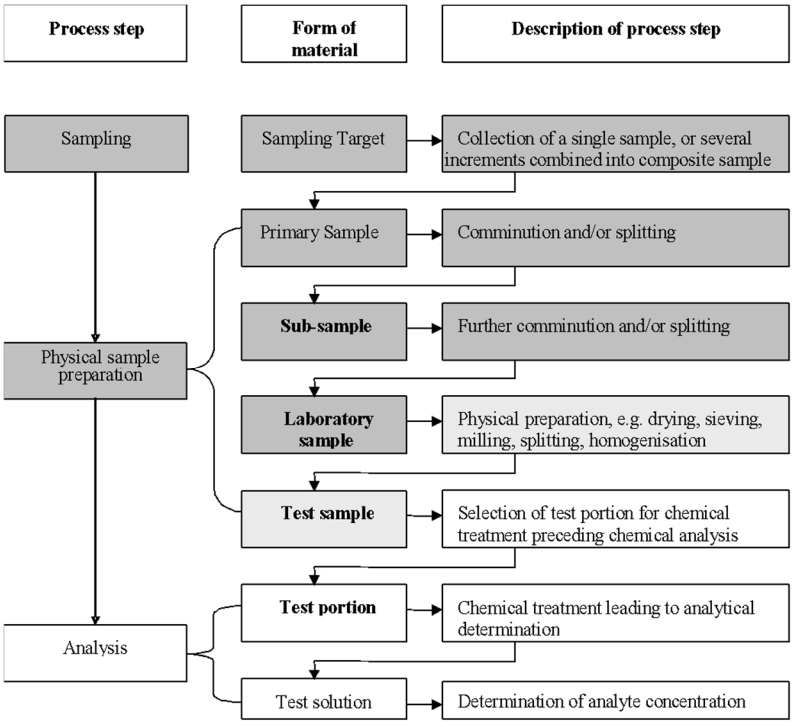

Figure 1 - Schematic diagram of the typical overall measurement process. The dark grey boxes show the sampling steps, the clear boxes the analytical steps, and the light grey boxes the physical preparation step which can be included in either category, depending on the experimental design used for validation [3]

1.7 What is a validated measurement procedure and how can its fitness be demonstrated quantitatively?

A sampling procedure cannot be validated in isolation, but it must be understood as one component of any measurement procedure. One important factor is that the validation of the measurement procedure provides quantitative evidence that the resultant measurement values meet the stated requirements of the measurement procedure. For example, if the uncertainty of the measurement values can be shown to be fit for purpose (e.g. measurement uncertainty sufficiently close to the target uncertainty). The validation will require some degree of replication to estimate the uncertainty of single measurement values. Routine application of the validated procedure, even for compliance testing, will not necessarily require replicate measurements.

A target value of measurement uncertainty, however set, can be used for the quantitative validation of the overall or *integrated* measurement procedure, thus including the sampling procedure. Validation of the overall *integrated* measurement procedure in one single operation is called the *simultaneous* approach. It may also be the case that the analytical procedure has already been validated in isolation,

either by single-laboratory validation in a competent laboratory, or *via* a collaborative trial whereby several laboratories are requested to strictly follow the same standard operating procedure. In this situation, the *sequential* approach can be followed, in which the sampling procedure is validated after the analytical (i.e. testing) procedure, to achieve validation of the overall measurement procedure. However, in this latter case, the validation of the analytical procedure should be reviewed and potentially revised in the context of this particular overall measurement process (See Section 2.1).

1.8 Terminology

For the purpose of this supplementary guide the definitions given in both Eurachem Guides [1, 3] apply generally, but otherwise an external source is cited. Terms are shown in bold on first use in the text.

The term 'measurement procedure' is used to include both the sampling procedure and the analytical procedure, with an occasional preceding adjective 'overall' to emphasise this situation. Analytical procedure is used (rather than the more traditional 'analytical method') to match the term 'measurement procedure' used and defined in VIM [6] (definition 2.6)[d], in preference to 'measurement method' that generally describes the technique selected [6]. The term 'measurement process' is used in the broader sense, equivalent to the VIM definition of 'measurement' (definition 2.1)[e], usually to emphasis the process rather than the result.

This document, and the Sampling Uncertainty Guide [3], both use 'estimation' of measurement uncertainty, based on estimated standard deviation. The QUAM guide [2] uses both 'estimation' and the term 'evaluation', that is used in GUM [7]. From the statistical point of view, there is a true (yet unknown) value of standard deviation and hence also of measurement uncertainty. Each measurement uncertainty value is therefore only an estimate with its own confidence interval [8].

The term 'concentration', when unqualified, should be understood as applying to any of the different measures of proportion or amount. When the text requires a restricted interpretation, 'concentration' is qualified (for example as 'amount of substance concentration') or replaced with a more specific term (for example, 'mass fraction').

[d] Detailed description of a measurement according to one or more measurement principles and to a given measurement method, based on a measurement model and including any calculation to obtain a measurement result.

[e] Process of experimentally obtaining one or more quantity values that can reasonably be attributed to a quantity.

2 Approaches

2.1 Including sampling procedures within integrated measurement procedures

ISO/IEC 17025:2017 [4] sets, as a general requirement, that '*the sampling method shall address the factors to be controlled to ensure the validity of subsequent testing results*'. In other words, the sampling procedure shall address all of the major factors which may influence the quality of a sample taken from a given sampling target (e.g., analyte **heterogeneity**, humidity, sunlight exposure, temperature, type of soil, amount/volume of sample, etc.), and any additional factors that arise during packing and transportation of samples. Each of those influencing factors shall be addressed and their possible effects on any compliance decision shall be evaluated in relation to any critical acceptance limits, using the overall measurement uncertainty reported alongside the measurement result for each primary sample.

The measurement uncertainty includes both random and systematic components, the latter of which are more difficult to estimate. The basic elements of sampling include deciding on how many samples to take (how many sample increments, if using a **composite sample** [3]), what mass (or volume) of sample to take (per increment) and where and when to take the sample(s) from the sampling target (according to any particular sampling procedure). All these aspects will influence the appropriateness of the sample and hence of the overall measurement procedure, and consequently the suitability of the resultant samples that are either brought to the laboratory or measured *in situ* by a portable device or sensor.

Thus, validation of sampling procedures must provide objective evidence that the requirements for a sampling procedure (whilst integrated within an overall measurement procedure, hence including testing, discussed below) for a given purpose are fulfilled, resulting in samples that are fit for measurement/testing (i.e. *appropriate samples* for that purpose). The overall requirement will be that the sampling is demonstrated quantitatively (i.e. using measurement uncertainty estimates) to be sufficiently representative of the parent sampling target [9].

The objective evidence is established by setting acceptance criteria for a number of performance characteristics (see Section 2.2) as appropriate in the given situation and demonstrating that these criteria have been fulfilled. A **representative sample** is defined as a '*sample resulting from a sampling procedure that can be expected to reflect adequately the properties of interest in the parent population'* [3]. A lack of representativeness is often seen as predominantly arising from the analyte heterogeneity in the sampling target. However, it must be emphasised in this context that it can also be affected by the technique applied by the individual **sampler**, and subsequent changes in the concentration of the analyte(s) of interest, between the moment of primary sampling and the application of the analytical procedure to the laboratory sample, due to factors such as inappropriate packaging, lack of control of important operating conditions during transportation, or operator mishandling.

There are exceptional cases where the primary sample is 100 % of the sampling target. In such a case, the sampling uncertainty component does not arise from analyte heterogeneity within the target, but arises from inappropriate packaging/transportation and variations in preparation processes such as sample storage, filtrations, etc. The latter contributions to uncertainty can sometimes be quantified using field blanks.

A sampling procedure can only be validated quantitatively as part of an overall measurement process, because it requires measurement values. An analytical procedure (or 'measurement method') can be considered in isolation, but this excludes the effects of all of the steps in the measurement procedure that occur prior to the analytical step. The degree of representativeness, reflected in the measurement uncertainty, can be seen as an important performance characteristic of the measurement process, once this process includes sampling. This is further elaborated in Section 2.2.

Some sampling procedures for specific applications are described in relevant legislative documents (e.g. Commission Regulation (EC) No 152/2009 laying down the methods of sampling and analysis for the official control of feed [10]). In the current regulatory practice, followed by the EC Directorate General responsible for the implementation of these policies, these sampling procedures are the result of extensive discussions among experts (from EU Member States, industry, academia, etc.) following dedicated proficiency testing (PT) rounds, collaborative trials or other research studies. These studies are organised to gather enough evidence for a particular analytical issue, e.g. demonstrating that laboratories are able to measure reliably at lower levels of a particular **measurand**, effectively allowing regulators to set a legal limit. The mentioned regulation states, '*sampling for the official control of feed, as regards the determination of constituents, additives and undesirable substances shall be carried out in accordance with the methods set out in Annex I'*. Moreover, '*preparation of samples for analysis and expression of results shall be carried out in accordance with the methods set out in Annex II'* [10].

2.2 Performance characteristics for measurement procedures that include sampling

It is understood that sampling must be done in accordance with a documented sampling procedure, including the extraction of a primary sample from the sampling target, containing (and if necessary, preserving it), transporting it to the laboratory and finally storing it. This procedure must be carried out according to the ISO/IEC 17025:2017 (Chapter 7.3) [4], following a **sampling plan** specifying the mass or volume of the primary sample, the number of sample increments (depending on the expected or known heterogeneity of the sampling target), place(s) and time, type and size of the container to transport the samples, conditions during sample transportation (except for *in situ* or on-site measurements) and any kind of preservation. These factors should be considered and optimised while designing each sampling procedure. The design of the sampling procedure depends on the characteristics of the sampling target (analyte(s) heterogeneity) and the purpose of the sampling, and hence on the purpose of the overall measurement process.

It is understood that some of the above-mentioned characteristics may be identified as the result of expert judgement (by competent samplers/inspectors who have the adequate experience/knowledge of the different sampling targets) that are subsequently validated. In routine use, minor deliberate deviations from written sampling procedures may be required due to the conditions under which samples are taken. When this is the case, a risk assessment should also be carried out to estimate the effects of these deviations on the quality of the laboratory sample, and eventually on the measurement result, and also on the final compliance decision. Any deviations should be clearly documented.

In addition to these deliberate deviations of the sampling procedure, a number of influencing factors (as mentioned above) may also have an impact on the suitability of the sample arriving at the testing laboratory, depending on the actual performance of the sampling procedure (and on the appropriateness of the sampling procedure). Looking at the overall measurement process (including sampling and analysis), the so-called performance characteristics for the overall measurement procedure must be applied.

The question is whether sampling, as part of the overall measurement process, requires specific attention to any of these performance characteristics (specifically including the degree of sample representativeness). The observations of each performance characteristic listed later in this section for each characteristic are based on informed judgement, but would require specifically designed experimentation to prove quantitatively.

The Eurachem FPAM Guide [1], prescribes experiments for the provision of objective evidence for every (relevant) performance characteristic of the measurement procedure. Evidence must be expressed in terms of specific quality requirements or performance characteristics.

Apart from the performance characteristic of 'selectivity', the evidence provided for the analytical methods is normally a quantitative measure in the form of a standard deviation based on the results of repeated measurements (carried out under specific measurement conditions). For the full measurement procedure that includes sampling, the objective evidence shall include all aspects from validation procedures already extensively and comprehensively described for analytical methods [1].

However, the inclusion of sampling in the overall measurement procedure requires that some additional steps are included. These have been described elsewhere [3]. The experimental design should include the estimation of measurement uncertainty, including the component arising from primary sampling (i.e. sampling uncertainty). This will typically require the replication of sampling (at least duplication) and each of these independent samples must be analysed using independent replicates (at least in duplicate) to enable the validation and thereby to ensure that appropriate samples are taken for subsequent analysis.

A particular challenge arises when two independent bodies/organizations handle sampling and testing separately. These bodies must work together, allowing the designs presented above to be executed.

Looking at the overall measurement procedure, its **precision** and **trueness** (bias), and consequently the measurement uncertainty associated with the measurement result, are the performance characteristics that may be expected to be most influential when sampling is included.

For the precision studies, it is important to include the effects of any relevant influencing factors from the sampling part of the process.

The validation statement, judging the fitness for purpose (based on objective evidence) of the overall measurement procedure (including sampling) shall be decided against pre-established criteria for the performance characteristics, for those which are affected while including sampling.

In case one or more of these criteria are not fulfilled, one option is to improve the sampling procedure, following a different sampling design or by increasing the number of sample increments in a composite sample. Alternatively, by improving the analytical procedure, whichever is more practical and cost-effective.

Measurement uncertainty has already been discussed as a criterion for judging the fitness for purpose of an overall measurement procedure. Measurement uncertainty has not always been included as a performance characteristic because it has been considered to be a property of the measurement result, not of the measurement procedure [1]. Applying the VIM definitions of a) measurement procedure and b) measurement result [6], the estimation of measurement uncertainty can be considered as part of the measurement procedure. Thus, measurement uncertainty can also be considered to be an additional performance characteristic of any measurement procedure [11].

The following performance characteristics, in addition to measurement uncertainty, should be considered individually in relation to evaluation of the suitability/appropriateness of a sampling procedure when integrated within an overall measurement procedure:

- **Analytical sensitivity**

'*The change in instrument response which corresponds to a change in the measured quantity (for example an analyte concentration), i.e. the gradient of the response curve'* [1]. This important performance characteristic shall be investigated during the development of a measurement procedure, as one of the primarily factors influencing its analytical, but not its sampling, contribution to the uncertainty;

- **Selectivity**

'*Extent to which the method can be used to determine particular analytes in mixtures or matrices without interferences from other components of similar behaviour'* [1]. This

performance characteristic is commonly assessed using an appropriate CRM (or an appropriate RM or internal quality control material). Sampling may be assumed generally not to influence this method performance characteristic;

- **Limit of detection, LOD / Limit of quantification, LOQ**

'*Lowest level of analyte that can be detected at a specified level of confidence'* (LOD), '*lowest level of analyte that can be quantitatively determined with acceptable performance'* (LOQ) [1]. Because an acceptable performance includes, necessarily, precision and trueness, i.e. measurement uncertainty, sampling may affect the LOD / LOQ. When sampling is a major component of measurement uncertainty, sampling may increase the LOD / LOQ of a measurement procedure (see working range). This can especially be the case if the sampling procedure introduces contamination. Sampling may also decrease the LOD. For example, if a large mass of stream sediment or soil (e.g. 50 kg) is taken for determination of gold concentration, then the effective LOD of the analytical procedure can be reduced by using field pre-concentration techniques (e.g. panning) to increase the number of grains of gold in the processed laboratory sample, and thereby potentially decreasing the sampling uncertainty component of measurement uncertainty;

- **Working interval (formally working range)**

'*Interval over which the method provides results with an acceptable associated measurement uncertainty*' [1]. Depending on the heterogeneity of the sampling target from which primary samples are taken, sampling may be a major component of the overall measurement uncertainty. In some cases, sampling may change the working interval of the method to keep the measurement uncertainty associated with a result within certain accepted limits (below or equal to the accepted target uncertainty), therefore, generally increasing the LOD / LOQ, and therefore reducing the working range, of the measurement procedure.

- **Trueness**

'Closeness of agreement between the average of an infinite number of replicate measured quantity values and a reference quantity value' [6]. The analytical component of the trueness can be estimated (as analytical bias) by using an appropriate (e.g. well matched) CRM/RM or a well-characterised internal quality control or PT item (with known quantity values). The sampling procedure should be designed to minimise sampling bias ([3], Section 10.2.4). However, residual or unsuspected **sampling bias** will also affect the bias (trueness) of the overall measurement result. This will not be apparent if the trueness is estimated with a CRM (thus ignoring the sampling component of the measurement uncertainty). To include sampling bias into the estimation of measurement bias, the use of either a **reference sampling target** (RST [12]) or the results of *Sampling Proficiency Testing* **(**SPT) [3] (or *Collaborative Trial in Sampling* **(**CTS) [13]) is required. In measurements of gaseous exhausts, a reference measurement is often performed annually in the same source, to check for bias.

- **Precision**

Sampling, if quantified as a measurement uncertainty component, certainly affects the precision of a measurement result by increasing, usually significantly, the observed variability, thus increasing both the imprecision of the measurement, and the measurement uncertainty associated with any measured value (see Appendix A, Examples A1 and A2);

- **Ruggedness**

'*Measure of its capacity to remain unaffected by small, but deliberate variations in method parameters'* [1]. It might be understood these variations are, mostly, related to the analytical

part of the measurement procedure. However, sampling may also significantly influence ruggedness if samplers do not follow a well-described sampling procedure (e.g., do not strictly respect the number and size of sample increments). Samplers often vary slightly in how they apply sampling procedures (deliberately or unknowingly, shown by video evidence), and this is often reflected in SPT performance scores. Sampling procedures for some (especially heterogeneous) sampling targets are not robust enough to reduce the effects of these differences on the resultant measurement values and their associated measurement uncertainty. The degree of representativeness of the sample (and hence the measurement) is quantified by the uncertainty associated with the measurement values (i.e. measurement uncertainty). This includes the contributions from sampling precision and sample bias (trueness component due to sampling). Measurement uncertainty includes the effects of both the heterogeneity of the analyte concentration in the sampling target, and also of the typical slight deviations that samplers make from the sampling procedure. The adding of more deliberate variations is probably not necessary when a CTS or SPT is used in validation (Step 3b below), but may be advisable if the Duplicate Method is used. Judging the fitness for purpose of the overall measurement (and hence including sampling) procedure, can be achieved by comparing the estimated overall measurement uncertainty against the target measurement uncertainty (See Section 2.3, and Appendix A, Examples A1 and A2).

2.3 Validation of measurement procedures that include sampling (VaMPIS) using the integrated approach, either sequential or simultaneous

2.3.1 Generalised approach

The validation of a sampling procedure within a measurement procedure needs to be quantitative to provide 'objective evidence' as required by ISO/IEC 17025:2017, (Chapter 7.2.2.1, including Note 1) [4]. This can be achieved in one of two ways; either following (a) a sequential or (b) a simultaneous approach [14].

In the sequential approach (Section 2.3.2) the analytical procedure/method has already been validated for the specified analyte and test material, whereby the performance characteristics of the measurement procedure [1] have been estimated (ignoring the sampling component).

In the simultaneous approach (Section 2.3.3), the analytical procedure/method is validated at the same time as the sampling procedure, therefore recognizing sampling as part of the overall measurement procedure. This latter approach is particularly relevant for *in situ* measurement procedures where the sampling and analytical steps are effectively inseparable (Sections, 2.3.3-6, 2.5 and the case study in Appendix A, Example A2). An overview of the validation is shown in Figure 2.

2.3.2 Sequential approach to VaMPIS

The sequential approach is only applicable if the selected analytical procedure (i.e. analytical method) is already validated for the specified analyte in the material comprising the sampling target. If this is not the case, the simultaneous approach needs to be followed (Section 2.3.3).

The descriptions of the 11 steps of this validation procedure (Figure 2) are:

Step 1: **Specify the measurand** of interest in terms of both the analyte and the sampling target (i.e. portion of material, at a particular time, that the (primary) sample is intended to represent, e.g. a lot, batch, or area). Check whether a target uncertainty for the overall measurement procedure has been externally specified by a regulator or customer (to inform Step 8).

Step 2: **Identify the detailed measurement procedure** proposed for the specified analyte, and the type of sampling target. This should include the sampling procedure (e.g. possibly using composite samples to reduce the sampling uncertainty component of the overall measurement uncertainty), any physical sample preparation (e.g. drying, sieving, filtration, milling, splitting, homogenisation, transportation), and a suitable analytical procedure that has been validated previously (Figure 1).

Step 3: Design the experiment to validate the measurement procedure (including sampling and analytical components)

Select the Duplicate Method [3] to be implemented by one sampler at each sampling target, or by more than one sampler if it is possible and more convenient for a particular organisation (e.g. testing laboratory). Sampling of different targets can be undertaken by two or more samplers, equally trained, providing that both halves of each duplicate sample pair are taken by the same sampler.

a. The Duplicate Method requires the selection of at least eight different sampling targets (which are selected as being typical of the specified type). Each sampling target is sampled twice using an independent or 'fresh' interpretation of the sampling procedure in a full, unbalanced or simplified balanced design (Figure 3, a, b or c). An example of this independence is the different, but equally likely, interpretation of the "W" design in Appendix A, Example A1.

b. A Collaborative Trial in Sampling (CTS) requires the use of at least one typical sampling target (but ideally several) and that each participant strictly follows the same sampling procedure. This target should be sampled in duplicate, and independently, by all of the different samplers, using a special balanced design (Figure 4 [13]). This option has the advantage of including the between-sampler bias (and potentially between-laboratory bias if participants make their own analyses) in the measurement uncertainty estimate and in the validation process. When the analyte heterogeneity within the sampling target is the overwhelming source of the measurement uncertainty, the extra contribution from the between-sampler bias may be negligible.

c. Within both of these experimental designs, physical sample preparation is usually included under the general heading of 'sampling'. If it has previously been identified as a potentially substantial source of measurement uncertainty, it can be estimated separately using an extended design also based upon the Duplicate Method [3, Fig D.1]. Alternatively, uncertainty from some physical sample preparation can be included under the heading of 'analysis' by making analyses of duplicated test samples split from the same laboratory sample, rather than on duplicated test portions from one test sample.

d. The Duplicate Method, SPT and CTS are examples of 'top-down' approaches, where the intention is to include all contributions to measurement uncertainty, without the requirement to estimate these contributions individually. The issue is then to decide which of the broad steps in the procedure need to have a separate estimate of their contribution to the measurement uncertainty. For example, the Duplicate Method automatically includes the measurement uncertainty component arising from physical sample preparation (and all steps before the selection of the **test portion**) under the label of 'sampling'. The sample preparation component would only need to be estimated separately if it was suspected of being so large as to require evidence that its subsequent reduction would be needed to achieve overall fitness for purpose [15].

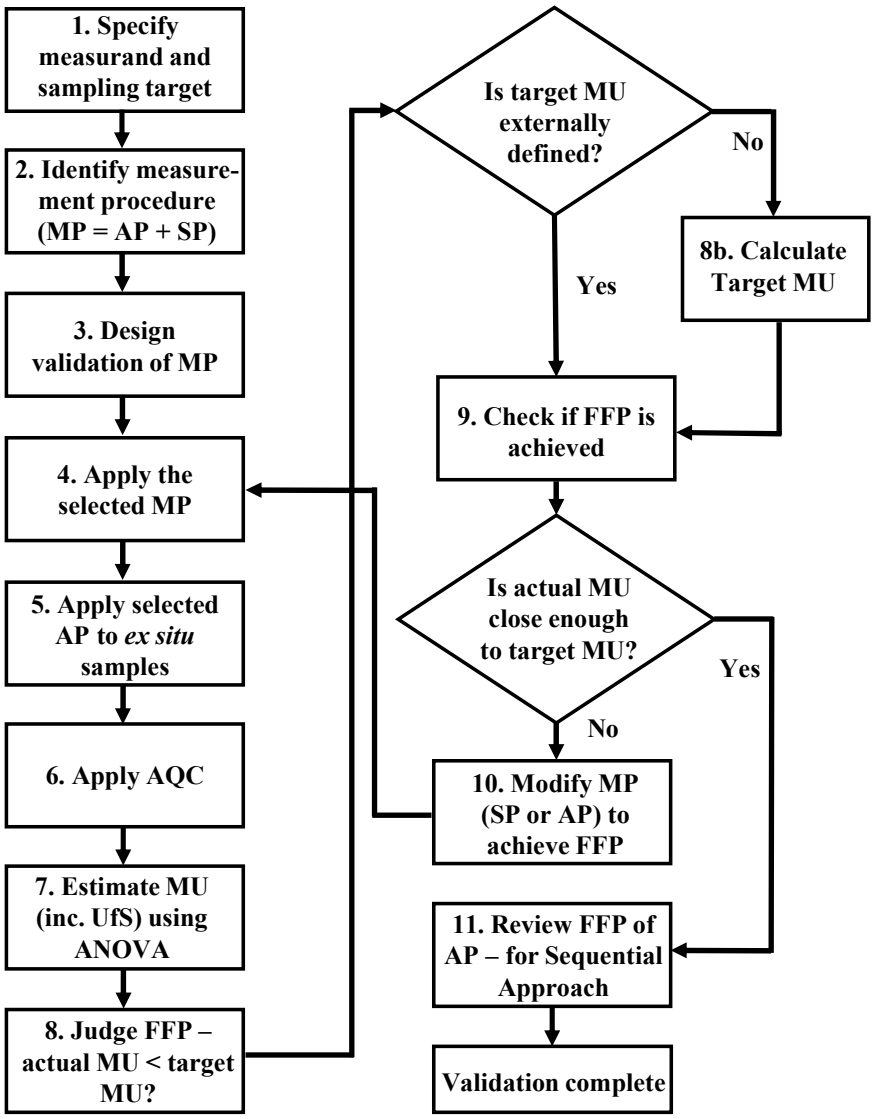

Figure 2 - Flowchart for quantitative and integrated VaMPIS (sequential or simultaneous). Shows 11 main steps by which a measurement procedure (MP) is assessed, with its components of a sampling procedure (SP) and an analytical procedure (AP). MU is measurement uncertainty, UfS is MU from sampling, FFP is fitness for purpose, AQC is analytical quality control. Reproduced from [14]

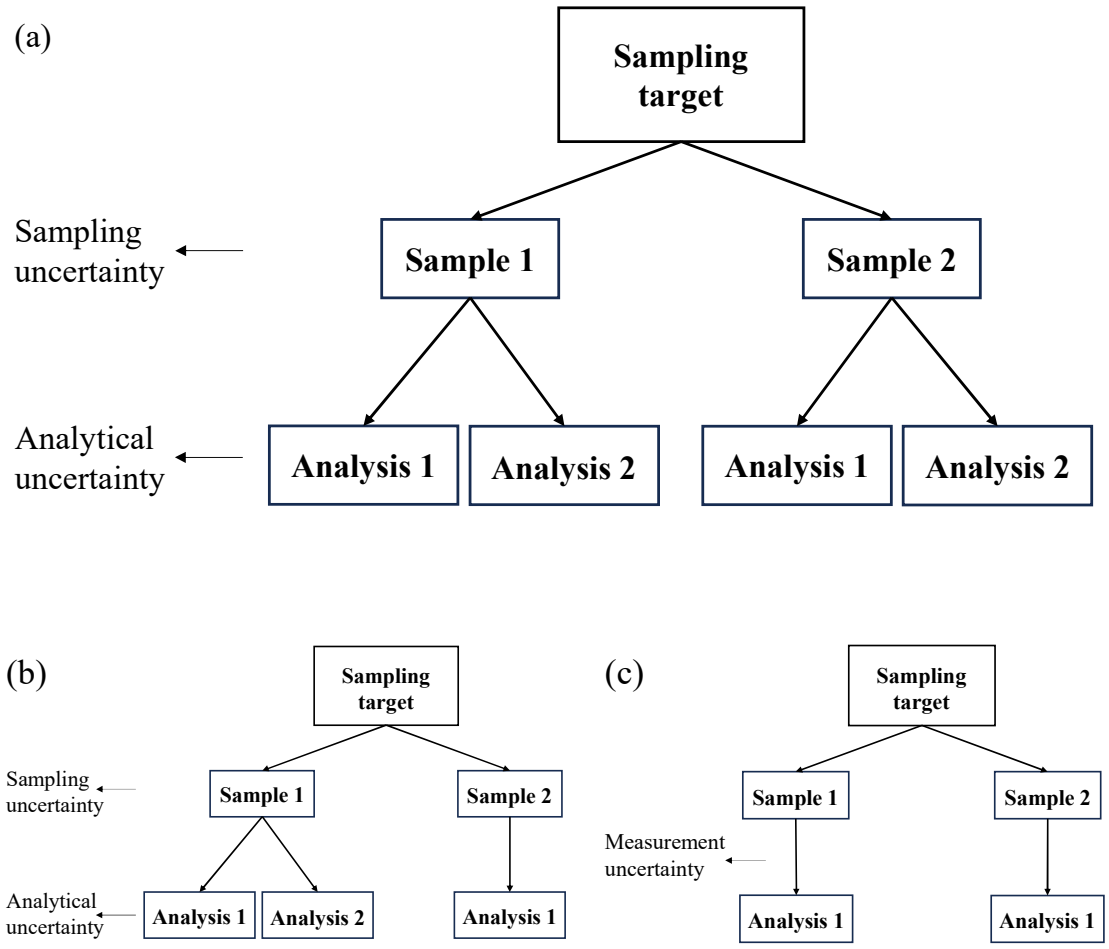

Figure 3 - The experimental designs used for the estimation of random component of measurement uncertainty (as repeatability) using the Duplicate Method: (a) Full two-stage nested balanced; (b) Unbalanced; (c) Simplified balanced [3]

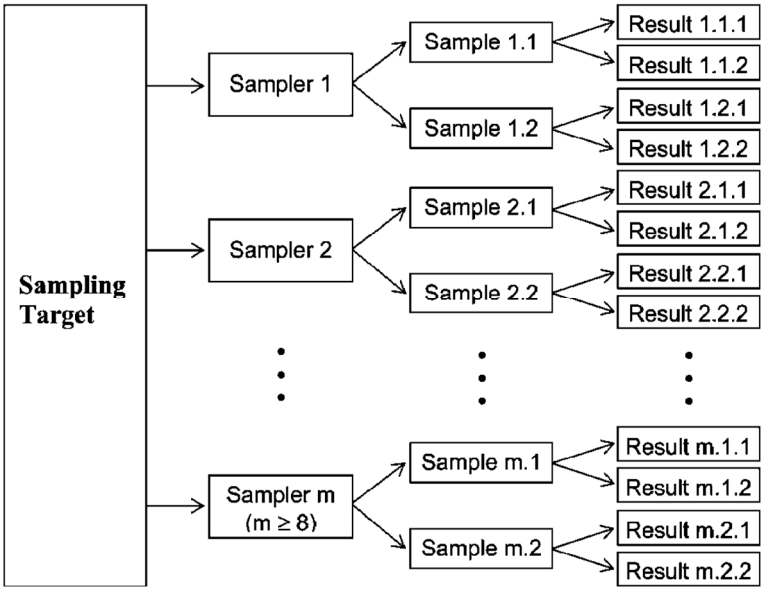

Figure 4 - Experimental design for inter-organisational sampling trial (SPT or CTS) [13]

e. Methods exist to estimate measurement uncertainty (including sampling uncertainty) as a function of concentration [16, 17], but require many more duplicate samples over a range of concentrations. Using the Duplicate Method, at concentration levels well above the limit of detection (i.e. > × 10) the relative uncertainty is fairly constant. Below that level, the absolute uncertainty is a more reliable estimate. The most important concentration level at which to estimate the measurement uncertainty is the regulatory threshold for conformity assessment, and the validation experiment can be designed to address that aim.

Step 4: Apply the selected measurement procedure, starting with sampling and physical sample preparation procedures for the *ex situ* option (including field blanks, and spiked test material, if feasible) to the sampling target(s).

Step 5: Apply the selected analytical procedure to all primary samples taken for the *ex situ* option. Make analyses of duplicated (and independent) test portions (or **test samples**) from both duplicated samples in the full balanced design (using either the Duplicate Method, Figure 3, or the CTS, Figure 4).

Step 6: Apply quality control (integrated, i.e. analytical and sampling**)** to all of the measurements in the routine way, including estimation of analytical bias and its uncertainty (e.g. apply corrections for analytical bias, reagent blanks and field blanks if necessary).

Step 7: Estimate the overall measurement uncertainty by applying ANOVA to the measurement values (where overall measurement uncertainty includes components arising from sampling and analysis). Include an estimate of analytical bias and its uncertainty in the measurement uncertainty estimate. Prior to estimation, investigate the frequency distribution of the measurement values to decide on the most appropriate type of ANOVA (classical or robust) and whether log-transformation (or other type of data transformation, if appropriate) is required. Further details and worked examples are available in Appendix A and elsewhere [3].

Step 8: Judging the fitness-for-purpose of the measurement results by comparing their measurement uncertainty estimates against a target uncertainty [3].

a. There may be an externally set target uncertainty specified by a regulator or customer (using guidance, e.g. [18]), against which the estimated measurement uncertainty can be compared to judge the fitness for purpose of the measurement results.

b. The target uncertainty can also be calculated for a particular analyte/sampling target (Step 8b), depending on the purpose of the measurements, using either the Percentage of Total Variance [3, Section 16.2] or the Optimised Uncertainty methodology. Details of the latter are given in Appendix B, also in the Eurachem Sampling Uncertainty guide [3, Section 16.3], and references [19, 20, 21].

c. The inputs required for the Optimised Uncertainty method include the experimental estimates of measurement uncertainty, and its individual components from sampling (u_{smp}) and analysis (u_{ana}). Also required are the costs of both the measurement components (sampling and analysis) and of the potential consequences of both false positive and/or false negative compliance decisions for the sampling target. The overall cost from both sources is minimised at the optimal level of measurement uncertainty, which can be used as the target uncertainty, and fulfils the definition of fitness for purpose (the theory of the Optimised Uncertainty method is

given in Appendix B, and worked examples of its application are given in Appendix A, Examples A1 and A2).

Step 9. Assess the extent to which fitness for purpose has been achieved by comparing the experimental measurement uncertainty against the known or estimated target uncertainty (e.g. Optimised Uncertainty) value.

a. If the experimental measurement uncertainty is sufficiently close to the target uncertainty, the overall measurement procedure can be said to be fit for its purpose (i.e. validated). The extent to which an arithmetic difference between the experimental and Target measurement uncertainty are significant can be judged statistically [18, 22], or financially by comparing the residual cost (i.e. expectation of loss) [19].

b. If, however, the experimental measurement uncertainty is substantially different from the target uncertainty, action needs to be taken to achieve fitness for purpose. In this case, a secondary part of the Optimised Uncertainty method can be used to decide whether it is more cost-effective to modify either the methods of sampling, or of chemical analysis, to achieve the optimal target uncertainty.

c. The breadth of applicability of this validation, and its potential to measurements on related crops, is discussed further in Section 2.6.

Step 10: Modifying the measurement procedure to achieve fitness for purpose (if required)

It may be more cost-effective to reduce the overall measurement uncertainty by modifying the sampling procedure to reduce the u_{smp} component, rather than to modify the analytical method to reduce its contribution. A typical approach to reducing the u_{smp} component of measurement uncertainty, and hence the overall measurement uncertainty, is to increase the number of increments used to constitute each primary composite sample. The factor by which to increase the number of increments to achieve fitness for purpose can be calculated theoretically, and subsequently tested experimentally [23]. A worked example is given in Appendix A, Example A1. It may also be that the dominant source of measurement uncertainty, classified as u_{smp} by the experimental design (Figure 3a), arises from the physical sampling preparation, rather than the sampling itself. In that case, a modified experimental design can be used to estimate and monitor any required reduction in this component [3, Figure D1 and 15]. If the measurement uncertainty is dominated by the analytical component, it is then most cost-effective to reduce the analytical measurement uncertainty. This can be addressed by considering the most limiting of the performance characteristics [1]. If that is the limit of detection, for example, then it may be sufficient to reduce that value in some way. Where both sampling and analysis are equally dominant sources of measurement uncertainty, the second part of the Optimised Uncertainty method can be used to judge the extent to which both may be reduced to achieve the target uncertainty most cost-effectively.

Step 11: Review fitness for purpose of the analytical procedure for the sequential approach

In the sequential approach, the analytical method is already validated for the specified analyte and test material. It is possible, however, that the measurement uncertainty reported as being fit for purpose in an isolated validation of an analytical procedure (u_{ana}) is statistically different from that estimated during an integrated validation including sampling. Such a comparison should allow for the confidence intervals of both measurement uncertainty estimates [22], where they are available.

If the measurement uncertainty of the analytical method estimated in isolation is significantly lower than that from the integrated approach, then the higher value from the integrated approach should be considered more realistic. This situation may arise because the measurement

uncertainty from the Duplicate Method includes a contribution from the heterogeneity of the routine test material (not the laboratory sample), which is usually higher than that of the much more homogeneous reference material which is traditionally used for this purpose in the isolated validation of the analytical procedure. However, it may also be useful to compare the quoted value of u_{ana} against the between-laboratory reproducibility from an inter-organisational trial (e.g. a CTS), to assess whether the former value is estimated reliably.

If the optimized level of u_{ana} from the Optimised Uncertainty calculation is significantly smaller than the experimentally estimated value, there may be a case for lowering the latter value, for example by lowering the analytical limit of detection. Alternatively, if the optimised value is significantly larger than the experimental value, then there may be a case for using a less precise (perhaps less expensive) analytical procedure that has a higher estimated measurement uncertainty. Worked examples of this general approach, including the Optimised Uncertainty methodology as well as the Duplicate Method, are included in Appendix A, Examples A1 and A2.

2.3.3 Simultaneous approach to VaMPIS

In the *simultaneous* and more integrated approach, the analytical procedure (i.e. method) is validated as part of the overall measurement process. The steps required (Figure 2) are generally the same as those for the sequential approach, with a few exceptions:

1. Where the sampling and analytical components of the measurement procedure are usually initially considered separately for the sequential approach, they can often be considered together for the more integrated simultaneous approach (especially for *in situ* measurements). The experimental design of the Duplicate Method can be applied in both approaches to provide an integrated assessment of both the whole measurement procedure and the relative contributions of its two main components;

2. Some performance characteristics of the analytical components of a measurement procedure are not easily assessed while following a simultaneous approach as described in the examples included in this Guide, e.g. analytical sensitivity, limit of detection / limit of quantification, working range and robustness. The broad suitability of these performance characteristics needs to be established in earlier experimental studies, although their effects in a particular application are reflected in the estimated measurement uncertainty that is used to judge the fitness for purpose of the overall measurement procedure.

   However, a simultaneous approach could also be followed in case of a measurement procedure applied with a set of replicates, on different days (replication within each day), using one or more operators (if relevant for a particular testing laboratory), on one or more instruments (if relevant), measuring a CRM or a well characterised internal QC material to assess trueness, and including different sampling targets (following the Duplicate Method as described).

   Indeed, many measurement procedures, due to their prolonged duration, won't allow the typically 32 (independent) analytical replicates (Figure 3a) to be measured under repeatability conditions of measurement. It might be necessary, therefore, to separate them, for example by making analyses for two sampling targets on each day (e.g. $n = 8$ spread over 4 days).

   This simultaneous approach would be able to assess the repeatability, intermediate precision (including day-to-day variability, within-laboratory operator variability and, if applicable, instrument variability), trueness/bias and the overall measurement uncertainty associated with the measurement result (while following this particular measurement procedure), including the sampling component.

This modified approach is generally not applicable to *in situ* measurement procedures in situations when there is the possibility of temporal variability in the analyte concentration within the sampling targets;

3. At Step 5 if a CTS is used, each participant conducts their own chemical analyses using the common analytical procedure that is to be validated, together with making measurements of a common matrix-matched reference material. The latter will enable between-laboratory analytical bias (and the extra measurement uncertainty it generates) to be quantified;

4. The analytical repeatability (from the ANOVA in Step 7) provides an estimate of that component of the measurement uncertainty (u_{ana}). This value can then be used as one parameter of the validation of the analytical part of the measurement procedure. Other performance characteristics (i.e. selectivity, limit of detection, limit of quantification, working range including linearity) may also need to be considered, and potentially adjusted, in this part of the validation to achieve the integrated target measurement uncertainty (including sampling and analysis);

5. The traditional target uncertainty (TU) is usually set for just the analytical component of the measurement uncertainty (i.e. TU_{ana}), but ideally it should be for the overall measurement that includes the sampling contribution (i.e. TU_{meas}). This can either be set externally by a regulatory body or agreed with a customer (for a particular analytical application) or by using an internal method such as the *Percentage of Total Variance* ([3]) or the *Optimised Uncertainty* method [3, Section 16.3] set up for the specific case under consideration.

    a. Generally, when TU_{ana} is not achieved initially, the information on all of the performance characteristics (listed in Section 2.2) can be used to select the most effective characteristics to adjust in order to achieve that target.

    b. If the sampling contributes the dominant contribution to the overall measurement uncertainty, and the cost of sampling is much cheaper than that of the chemical analysis, then the experimental estimate of u_{ana} may be lower than that required to meet the overall target measurement uncertainty. In that case it may be that a less expensive analytical method is sufficient for this purpose (e.g. using a shorter counting period on a spectrometer's detector). Conversely, in the opposite circumstances, a reduction in u_{ana} may be required (e.g. use of an internal standard in spectrometry, or selection of a different analytical technique with a lower limit of detection);

6. The *simultaneous* and more integrated approach is particularly appropriate for *in situ* measurement procedures, where no physical sample is extracted from the sampling target. In this case, the two processes of sampling and analysis are effectively inseparable and an integrated approach to the validation is essential. The same two approaches can be used (Step 3), but the '**duplicate samples**' are taken by relocating the *in situ* measurement device with the spatial and temporal ambiguity implicit in the measurement procedure (Section 2.4, with worked example in Appendix A, Example A2).

Irrespective of the approach taken to validation, ongoing sampling and analytical QC will be required to monitor the measurement uncertainty (and its components u_{smp} and u_{ana}) to see if they vary significantly from the value established during the validation, perhaps using 'repeatability limit' [1, Section 6.6.3]. This will be discussed in more detail in Section 3.

2.4 *In situ* methods

2.4.1 Validation of *in situ* methods using the integrated approach

Measurements that are made *in situ* do not require the extraction of a physical sample but involve placing some sort of measurement or sensing device at the original location of the sampling target.

The measurement method begins with placing the measurement device (e.g. a sensor) in a particular place (touching or very close to a given sampling target) and at a specified time, to best represent the stated sampling target. This type of sampling results in a 'virtual sample' with dimensions of space and mass that are usually determined by the measurement technique, and may not be accurately known by the operative.

The device then analyses a portion of the sampling target. The portion analysed can therefore be considered to be an '*in situ*' sample, 'taken' but not removed or extracted from its original position in space and time. Examples of *in situ* methods include sensors measuring (a) different gases within an industrial stack or chimney, (b) contaminants in a flow of water (both over a specified period of time), or (c) trace elements in topsoil across a specified area of land.

An example of (c), is Portable X-ray Fluorescence Spectrometry (pXRF) for which the depth and hence mass of the *in situ* sample will vary between different analyte elements (e.g. 1 to 320 mg) depending on the analysing depth (e.g. critical x-ray penetration depth) in that test material [24].

The 'analytical' part of the measurement procedure in all cases is effectively inseparable from the 'sampling' part. All parts of the measurement procedure are undertaken by the same person (or machine) with little of the clear divisions between sampling and analysis that are usually present for *ex situ* measurements (where an extracted sample is analysed in a remote laboratory).

Moreover, the virtual sample is not prepared/processed (e.g. dried and/or homogenised) as is usual in *ex situ* procedures or methods. In the case of a virtual sample the original heterogeneity of the analyte concentration is not decreased, as would be the case for grinding of an *ex situ* sample, and it is therefore often a cause of an increased contribution from the u_{sna} component of the overall measurement uncertainty. Additionally, so called portable or handheld devices often do not possess the high level of measurement performance of *ex situ* instrumentation, such as resolution, spectral range, dynamic range, etc.

The validation of an *in situ* measurement procedure ideally happens predominantly at the location of a typical sampling target (e.g. 'in the field'). However, it has often been considered more convenient to test the *in situ* analytical device (e.g. a sensor) in a laboratory, but this test will then be performed under relatively ideal conditions that will not replicate the conditions in which the routine measurements are made. Any value of measurement uncertainty estimated by a manufacturer of an *in situ* measurement device is generally more likely to be an estimate of u_{ana} based on within-laboratory repeatability, and exclude analytical bias, between-laboratory bias, and often most importantly, ignoring the u_{smp} component of the overall measurement uncertainty.

In particular, laboratory-based tests do not include the interaction of the measurement device with a real-world sampling target. Validation with a certified reference material (CRM), for example, would omit the taking of an *in situ* sample from a typical sampling target, which is often substantially heterogeneous (both laterally and vertically).

The effective integration of the sampling and analytical steps in an *in situ* measurement procedure means that an integrated approach for validation of the overall measurement procedure (Section 2.3) is the preferable option. Methods already described for this purpose for *ex situ* are broadly applicable, with minor adaptations, but some additional steps are also required, as described below and in Sections 2.4.2 – Section 2.4.4:

> *Water Quality Management.* EN 17075 [25] describes the requirements and protocols to assess the performances of continuous measuring devices used either for discrete

measurement (portable devices) or in a fixed position for continuously measuring water quality. After assessing the performances in controlled conditions (metrological performance characteristics and any relevant factors that may influence the measurement) a 3-months field trial is recommended to check that the device operates in real conditions with the same performances. EN 17075 does not use the Duplicate Method with ANOVA. Instead, it requires 24 paired measurements (devices and reference method) during the field trial. The 90th percentile of the differences (in absolute value to compensate for under / over estimation) is then calculated and compared to the measurement uncertainty value estimated under controlled conditions;

Air Quality Measurement. ISO 20988:2007 [26] provides comprehensive guidance and specific statistical procedures for uncertainty estimation including measurements of ambient air, stationary source emissions, indoor air, workplace atmospheres and meteorology. It applies the general recommendations of the GUM [7] to boundary conditions met in air quality measurement. The boundary conditions considered include measurands varying rapidly in time, as well as the presence of bias in a series of observations obtained under conditions of intended use of methods of air quality measurement. The methods of measurement considered comprise:

- methods corrected for systematic effects by repeated observation of reference materials;
- methods calibrated by paired measurement with a reference method;
- methods not corrected for systematic effects because they are unbiased by design;
- methods not corrected for systematic effects in intended use deliberately taking into account a bias.

Experimental data for uncertainty estimation can be provided either by a single experimental design in a direct approach or by a combination of different experimental designs in an indirect approach.

2.4.2 Estimating the repeatability using the Duplicate Method

When applying the Duplicate Method (Section 2.3.2 Step 3), a duplicated 'sample' must be taken by repositioning the *in situ* measurement/sensing device using a reinterpretation of the instructions for locating the device in space and/or time.

The full balanced design (Figure 3a) requires two analyses to be made for both duplicate 'sample' positions. Alternatively, only one analysis can be made for both 'sample' duplicates in a simplified design (Figure 3c). In this latter case an external estimate of the u_{ana} component will be needed to enable the separation of the sampling component (u_{smp}) from the combined measurement uncertainty (u_{meas}) that has been estimated using ANOVA (see Example A2). The time saved by using this simplified balanced design can be used to measure a greater number of sampling targets, and thus reduce the confidence interval on the estimates of measurement uncertainty.

2.4.3 Estimation of analytical bias

The estimation of analytical bias for a method used for an *ex situ* procedure (Figure 2, Step 6, e.g. using matrix matched CRMs) can also be applied to an *in situ* measurement device, but these usually underestimate the bias of *in situ* measurements on real sampling targets. Causes for this underestimation include differences between the composition and properties of the CRMs and the test material (and the sampling target), due to issues such as moisture, grain size, heterogeneity and surface roughness [27].

Such components of the systematic effects can be estimated by including in the validation a comparison against *ex situ* measurements made on physical samples taken at the same locations as those used for the *in situ* measurements. A statistical model can then be constructed to

describe the relationship between the *in situ* and the *ex situ* measurement values, whilst allowing for measurement uncertainty of both types of measurement procedure.

The resultant model gives an estimate of the bias between both sets of measurements which has two components (each with their own uncertainty): a fixed value called the translational bias (given by the intercept coefficient), and a proportional component called the rotational component (given by the slope coefficient, Figure 5). A worked example of this approach is described in Appendix A, Example A2.

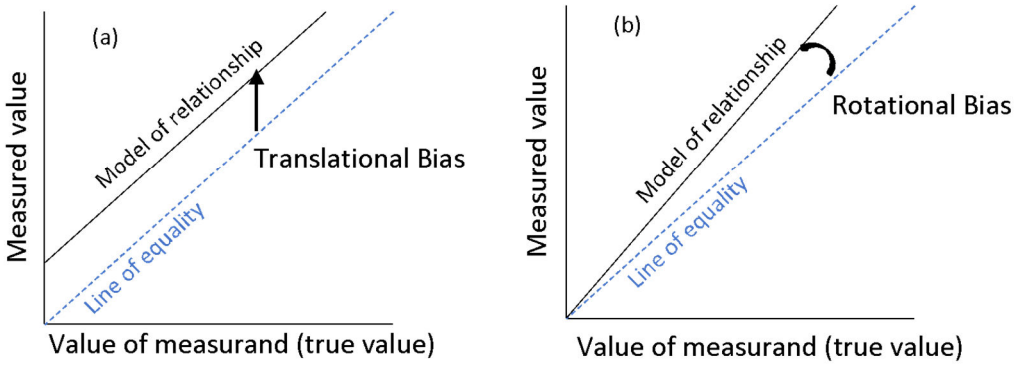

Figure 5 - Schematic representation of the two components of measurement bias (e.g. for an *in situ* method) as a function of concentration. Bias is where all measured values are different from the value of the measurand (i.e. true or accepted reference value) by either (a) a fixed concentration for translational bias, or (b) a proportion of the concentration for rotational bias. The value of the measurand can be represented by either the certified value of a series of matched CRMs, or measured values made by a second 'reference' method (such as an *ex situ* method)

2.4.4 Matching the measurand and the sampling target

Before comparing an *in situ* against an *ex situ* measurement, it is important to match the specification of both the measurand and the sampling target. The value of the measurand (defined as the 'quantity intended to be measured'[6]) is equivalent to the true value of the analyte concentration in the sampling target. In the case of lead in soil (Appendix A, Example A2) the *ex situ* method (ICP-AES after acid digestion) measures the total lead concentration in dried, sieved and ground top soil.

By contrast, the *in situ* method (pXRF) also measures total lead, but in unprocessed soil that still contains moisture, biota (e.g. plant roots), coarse-grained material not within the definition of 'soil' (e.g. stones > 2 mm diameter) and pore spaces. It may be impossible to match the two measurands perfectly, but they can be identified and sometimes the matching improved. For the soil example, if the measurand is defined as the concentration reported on an oven dried basis, this could be achieved by measuring the field moisture of the soil (ideally with a probe at each location) and correcting for it. A substantial reduction in measurement bias has been reported using moisture correction [27].

The sampling target also needs to be matched between the two measurement techniques as far as possible. For the soil example, the depth of top soil traditionally sampled for *ex situ* analysis is 150 mm, whereas the critical penetration depth of *in situ* pXRF is less than 1.5 mm for lead [27]. This mis-match will be important if the analyte concentration is shown to vary with depth, for example when revealed by using the pXRF to measure the lead concentration over the core's depth of 150 mm [27].

Alternatively, the depth of the *ex situ* sample can be reduced to 1 mm, but this will produce a primary sample of low mass (e.g. ~ 0.5 g) and consequently higher sampling uncertainty component and larger overall measurement uncertainty for the *ex situ* measurement.

2.5 On-site methods

On-site methods involve taking samples and analysing them 'on-site'. A physical sample is extracted from the sampling target, usually prepared in some way, and then analysed nearby instead of being transported to a remote laboratory. In essence, these are very similar to traditional *ex situ* lab measurements, except for the place in which the analysis is conducted, which is on-site rather than in a remote laboratory.

For the purposes of validation, the procedure is therefore basically identical to that described for *ex situ* measurement procedures that include sampling (Section 2.3). The only difference is that the performance characteristics (including measurement uncertainty) need to be quantified with the measurements made on site, rather than in the lab.

The on-site conditions are usually more variable and less controllable than those in the lab, and different instrumentation and reduced staff supervision may be additional factors that affect the measurement uncertainty. This would be expected to produce measurements with different characteristics (e.g. u_{ana} component) that need to be combined with the u_{smp} component to estimate the overall measurement uncertainty, as has already been described for the *ex situ* situation.

2.6 Validation reports and validation statements that include sampling

The general requirements for preparation of a validation report are described elsewhere [1, Section 5.3]. That description needs to be broadened to include sufficient detail of all the steps in the measurement procedure, including primary sampling, preservation and physical preparation of the sample prior to arriving at the laboratory, in the case of *ex situ* measurements. For *in situ* measurements, the report needs to explain how the measurement device was located at the sampling target in the validated procedure, and how much flexibility was allowed for in the interpretation of that procedure. In all cases the scope of the validation needs be made explicit in terms of the analyte and sampling target for which it is applicable. For example, in the case of *ex situ* nitrate determination in field-grown lettuce batches (Appendix A, Example A1), the validation would apply to similar batches of many thousand heads of lettuce, but not necessarily to other crops, or to lettuce heads in a retail setting. For *in situ* measurement of lead in top soil (Appendix A, Example A2), the validation for geochemical mapping is applicable for that particular test site, but would not be automatically transferable to other contaminated land sites without QC evidence to support that wider application.

3 Follow up after validation

3.1 Ongoing validation

To ensure the fitness for purpose of measurement results in routine operation, there is a very important relationship between validation and the ongoing quality control of measurement procedures that include sampling.

Validation of a measurement procedure provides objective evidence that the selected measurement procedure is fit for a specific purpose. Furthermore, the validation will provide information about which influencing factors are critical and should be described in the measurement procedure, including testing and sampling, to ensure the validity of the final measurement results. This will enable the preparation of an appropriate quality control plan that will cover all the factors that need to be monitored. However, validation is a one-time event and cannot ensure that the method is fit for purpose for routine daily use as conditions during routine sampling and testing may differ from those during the validation [1, and 3, Section 13.1]. Sometimes, some changes may be made such as the introduction of new equipment or new personnel or method improvement, for example in case one needs to extend the (analytical) scope of the method (including an extra matrix) or extend the working range to a lower (or higher) analyte content. In these cases, the influence of such changes should be determined.

Ongoing validation ensures that the measurement procedure remains valid during routine use and thus fit for the intended purpose of the measurement values. It includes both routine monitoring of the measurement procedure (i.e. QC) and evaluation of the procedure's performance after changes have been made, in order to determine whether the measurement procedure is still fit for purpose.

3.2 Quality Control as an integrated part of an ongoing validation

Quality Control (QC), should be carried out throughout the overall measurement procedure. Ideally, following an *integrated measurement QC approach* (IMQC, thus including sampling). QC measures are well established and routinely applied in laboratories. Replicated (e.g. duplicated) sampling shall be included, in addition to replicated analyses, in order to check whether the overall measurement uncertainty estimated during the initial validation is still applicable. Replicate samples can be used in an integrated approach to simultaneously check the measurement uncertainty component due to sampling (u_{smp}) and the analytical (u_{ana}) component, either as u_{meas}, or individually, depending on the experimental design used. Other possible QC materials include field blank samples to test for possible sample contamination, and spiked samples to check the stability of the analyte in the sample during transportation and storage. When the analytical procedure includes extensive sample preparation, the field blank sample and the spiked samples used for QC can also be used as control samples to evaluate method bias in an integrated approach. To ensure the fitness for purpose of routine analysis, a suitable *integrated measurement quality control plan* (IMQCP), which includes both analytical and sampling components, should be prepared and followed. The QC frequency will depend on the risk assessment, and when the risk is high, a higher frequency of quality control is recommended.

Whenever a change in either the sampling or analytical part of the measurement procedure is made, it is necessary to check whether this change will affect its initial performance. In an integrated approach, this can be done by re-calculating the u_{meas} value after applying a simplified balanced design (Section 2.3.2, Figure 3c) including the change carried out, and comparing it against the u_{meas} estimated during the initial validation (using the integrated approach). If the new u_{meas} value (again with integrated approach) is not significantly different from the

measurement uncertainty estimated during the initial validation [22], the method can be considered as still fit for purpose and the original u_{meas} value should be reported. Otherwise, the overall procedure should be revalidated.

Requirements of the ongoing validation parameters should be the same as, or close to, those of the initial validation. Ongoing validation should be carried out according to an approved procedure and should be well documented. All results should be analysed by statistical tools and clear conclusions should then be stated.

3.3 Monitoring of the measurement process over time through establishment of appropriate IMQC systems

The purpose of an appropriate Measurement Quality Control (QC) is to demonstrate that the overall measurement procedure is adequately controlled and fit for its intended purpose. The QC programme must cover all of the steps taken, including, for example, the physical preparation of the primary samples, to ensure that the reported measurement results are fit for purpose.

In an integrated approach the IMQC programme must also include features that will cover all steps of the sampling procedure: preparation of containers, equipment checking, sampling, transport as well as the competence of all personnel performing the sampling. The sampling part of an IMQC programme includes documented evidence that:

- personnel performing sampling are competent and professionally trained, and the ongoing evaluation of the sampler's competence (Section 3.5) is carried out according to the pre-planned design and frequency;
- procedure applied to the collection and handling of the samples is appropriate and validated;
- sampling equipment is regularly maintained and calibrated where possible;
- complete and secure sampling documentation is sufficient to ensure the logistical traceability and unique identification of the sample.

For sampling, the IMQC features include adequate monitoring and control of the sources of sampling uncertainty (i.e. as a component of the overall measurement uncertainty) such as analyte heterogeneity within the sampling target, instability, and possible contamination of the primary sample. IMQC procedures need to enable the efficient detection of sampling uncertainty significantly above the level that was estimated during validation, and provide a way to reject invalid measurement results arising from sampling that are not fit for purpose.

The objectives of an integrated QC (IMQC) approach are:

- monitoring of the overall measurement uncertainty (u_{meas}) and its components (u_{smp} and u_{ana}) *where possible*, to confirm that they do not deviate significantly from the values established during validation;
- monitoring sample contamination during sample taking and handling (where possible);
- monitoring the stability of primary samples during transportation and storage.

If validation has shown that there is no risk of sample instability, this aspect can be omitted, but this should be documented in the validation report.

Internal quality control features can include:

- use of reference sampling targets (if available), reference materials or quality control materials;
- use of alternative instrumentation that has been calibrated to provide traceable results;
- functional check(s) of equipment;

- use of check or working test materials with control charts, where applicable;
- intermediate checks on measuring equipment;
- replicate tests or calibrations using the same or different methods;
- retesting of retained test materials;
- correlation of results for different characteristics of a test material;
- review of reported results;
- within-laboratory comparisons;
- testing of blind test material(s).

External quality control measures can include:

- interlaboratory (or inter organisational) comparisons such as Collaborative Trial in Sampling (CTS) and Sampling Proficiency Testing (SPT) (Section 3.4).

The data obtained from IMQC should be recorded in such a way that trends can be detected and, where possible, statistical techniques should be applied to analyse the results [3], and, if applicable, improve the laboratory's performance. If it is determined that the IMQC results are outside the predefined criteria, appropriate actions should be taken to prevent the reporting of incorrect results, possibly based on risk assessment [4].

In the case when a laboratory/organisation performs only sampling, without making any measurements, it is not possible for that organisation to perform validation (Section 2.1) nor to apply IMQC procedures that provide quantitative results, such as monitoring of measurement uncertainty, sample contamination or sample stability. The best option in this situation is for the organisation responsible for sampling to collaborate with the testing laboratory in order to undertake both the integrated validation and the subsequent IMQC.

For sampling accreditation according to the ISO/IEC 17025 [4] standard, which describes the general requirements for the competence of testing and calibration laboratories, a laboratory must confirm all requirements related to all laboratory activities.

The quality control program will be more effective if it includes more than one sampler to see changes in sampling, heterogeneity and precision, field inspection and equipment validation.

In the case of non-compliance, all deviations should be investigated.

3.4 Participation in PT schemes (including sampling)

The benefits to analytical data quality from laboratories taking part in Proficiency Testing (PT) schemes is now well established. Extending this principle to the overall measurement procedure, by including sampling and sample preparation procedures, was suggested in theory in 1995 [28], and demonstrated in practice later that year.

Sampling Proficiency Testing (SPT) schemes have since been adopted by several application sectors. However, in some cases these have been limited to the primary sampling steps.

Most SPTs can effectively be considered as *Measurement Proficiency Testing* (MPT) schemes. This is because an MPT requires replication of the overall measurement procedure. The participants of an MPT (often called an SPT) are not only required to take the primary samples, but also conduct the physical sample preparation, as well as the chemical analysis, even if the latter is delegated to another organisation.

The results from an SPT provide quantitative evidence. They enable comparisons to be made between several different implementations of a given sampling procedure, where each of these implementations is performed by a different sampler. Importantly, each sampler in an SPT uses the same combination of analyte and sampling target. The information provided by an SPT can be useful for: a) the training of new samplers; b) ongoing monitoring; c) improving the performance of more experienced samplers.

Furthermore, SPT results can provide traceable quantitative evidence, which can be used either for the certification of samplers, or the accreditation of sampling organisations (e.g. to ISO/IEC 17025 [4]). SPT results can also be used to make a more rigorous estimate of the measurement

uncertainty component arising from sampling (sampling uncertainty, u_{smp}) that includes the systematic effects between different samplers. This value can then be compared against the sampling uncertainty value that was estimated at the time of validation (following an integrated approach). If it is found to be significantly larger, then this would suggest that either a) the sampling procedure needs to be reviewed and tested in a further SPT, or b) the validation needs to be reviewed, and repeated or re-evaluated with the new value of sampling uncertainty component (u_{smp}) if necessary.

Traditionally, PT schemes have utilized very homogenous PT items. This means these materials need no further processing or subsampling after a single test portion is taken. Consequently, estimates of measurement uncertainty are likely to be underestimated, because they ignore other components of measurement uncertainty that would usually arise from the subsampling and processing of a field sample, e.g. chopping, drying and homogenization. These limitations can be overcome if the u_{smp} component of the overall measurement uncertainty is estimated from SPT results that a) include replication of all of the procedures that would be used for a real field sample; b) are obtained from measurements made on realistically heterogeneous sampling targets, ideally in field conditions.

Separation of analytical PTs and sampling PTs can result in the exclusion of effects that arise from the overall measurement procedure. This is why integrated MPTs are to be preferred. Once the overall measurement procedure has been validated, using the steps described in this document, participation in an MPT is recommended in order to monitor the proficiency of the practical implementation of the overall measurement procedure.

3.5 Ongoing evaluation of qualifications of samplers

A sampler should document their competence with respect to sampling in general, and to sampling of those matrices covered by the scope of the sampling procedure. Competence could be obtained and maintained by participation in theoretical and practical and/or specialist courses, seminars, by experience, by on-the-job training under instruction of an experienced sampler, and by participation in SPTs, where available (Section 3.4), and sampler certification where applicable [29, 30].

The technical skills required include understanding of at least: the purpose of sampling; sampling as part of the measurement process and its contribution to the measurement uncertainty; principles of sampling techniques; limitations of the equipment; equipment control; calibration and maintenance; practical sampling and subsampling; principles of sample storage and transport including proper sample container materials and risks of cross-contamination; and finally the influence of critical factors.

The quality of the sampling activities and consequently the performance and qualifications of samplers are monitored (IMQC). The performance and qualifications should be evaluated routinely by defining an on-going specific surveillance process. The surveillance should ideally be based on quantitative, but also descriptive performance data obtained during the sampling activity, including quality control data, on-going regular training, and compliance summary. The latter descriptive elements are particularly relevant in the case of sampling activities not performed by a testing laboratory. If necessary, the performance assessment can also be based on spot checks in the form of interviews and/or inspection.

The evaluation frequency of qualifications of samplers can be undertaken yearly as a minimum for best practise. However, the frequency should consider relevant factors such as changes to the relevant standards and regulatory requirements, risks resulting from sampling by an insufficiently trained or unfamiliar operator/officer, and ongoing changes in technology. The amount of sampling work required in order to maintain sufficient experience levels depends upon the role of the sampler.

The purpose of ongoing evaluation is to demonstrate and validate the knowledge and competence of samplers even in cases of unexpected or unusual circumstances.

4 Management Issues

4.1 Management of the whole measurement process

4.1.1 General

The measurement process consists of several steps (summarized in Figure 1). During the preparatory phase of the activities, it is necessary to define among the interested parties a person responsible for the overall measurement process. Ideally, the laboratory manager should be responsible for the overall measurement process. As the sub-responsibility for the sampling and the appropriateness of samples can vary depending on circumstances, a general framework (Table 1a and 1b) can be used to specify areas of responsibility in advance, and is shown here for one particular example. To ensure the integrity of the samples, all steps of the sampling procedure should be identified and documented and the so-called 'chain of custody' needs to be established (for logistical rather than metrological traceability). As sampling is (mostly) based on one or more people being responsible for the various steps in that process, it is equally important that the role of each person is clearly defined and determined before starting the process (and also have them clearly identified in the chain of custody).

Depending on the overall setup and purpose of the measurement process (and as such the sampling part of it), different people may be involved, depending on the particular situation (listed in the columns of Table 1).

The following subsections will identify several responsibilities undertaken by the different people involved, based on their role and their background training and experience for fulfilling that role.

4.1.2 Establishing the sampling procedure

As the sampling procedure sets the framework for the actual sampling to be done, it is crucial that those taking responsibility for establishing such procedures and plans are fully aware of its importance and the purpose of sampling within an overall measurement procedure. The designers of the sampling procedure must be people with knowledge and experience with regard to the form and conditions of the sampling target, although that should also involve the customer. The input from experts in the field will often be needed (Table 1).

When a sampling procedure has been established and validated externally (e.g. by a regulator or similar) the sampling / measurement team has the responsibility to follow this procedure and apply IMQC as appropriate.

Furthermore, the laboratory doing the subsequent analysis of the samples should be consulted for any specific requirements. Sampler organisations and testing laboratories should collaborate to achieve reliable testing results.

4.1.3 Primary sampling

Prior to scheduling the sampling, all relevant information shall be shared between the testing laboratory, sampling staff and the ultimate customer/client. The information must be stated clearly in a Sampling Procedure document which will have the purpose of stating the correct information once in the field (i.e. at the **sampling location**).

Table 1 - Example of measurement responsibility* framework in the routine measurement phase (after validation): (a) overall responsibility of laboratory manager, sampling performed by laboratory, (b) overall responsibility of regulatory/inspection body

(a) External stakeholders	Client						
Internal Stakeholder		Sampling manager	Laboratory sampler	Internal courier	Sample reception	Laboratory	Laboratory manager
Overall responsible Measurement phase							X
Sampling request and information gathering	X	X					
Sampling design (or selection of appropriate sampling procedure (ideally previously validated)		X	X				X
Primary sampling (including *in situ* analysis)	X		X				
Sample shipping			X	X	X		
Verification (suitability) and receipt of samples			X		X		
Laboratory analysis (including sample preparation)						X	X
Quality of the final measurement result	X						**X**

(b) External stakeholders	Client			Parcel courier	External laboratory
Internal Stakeholder		Inspection body	Professional sampler		
Overall responsible Measurement phase		X			
Sampling request and information gathering	X	X			
Sampling design (or selection of appropriate sampling procedure (ideally previously validated)		X	X		X
Primary sampling (including *in situ* measurement)	X		X		
Sample shipping			X	X	X
Verification (suitability) and receipt of samples			X		X
Laboratory analysis (including sample preparation and validation of analytical process)					X
Quality of the final measurement result	X	X (*)			

(*) *there needs to be ongoing communication and cooperation between the Regulatory/Inspection body and the laboratory manager, who is responsible for just the analytical component as well as to provide evidence of the sampling quality. When the validation of a measurement process involves different sampling targets (or measurands), it is essential that the validation report includes the minimum requirements that are needed for the validation to be effective (e.g. personnel qualifications/job titles/responsibilities).*

The minimum information to be stated within the sampling procedure are:

- scope of the measurement, including:
    - the extent of the sampling target (in space and/or time);
    - the definition of the measurand (including units of measurement, and whether to be reported on an original/fresh or dried mass basis);
    - the relevant applicable legislation and regulatory limits for compliance assessment (if applicable).
- analyte(s) or parameters to be measured in the laboratory sample;
- any measurements or analyses to be taken in the field (either on site, or *in- situ*, i.e. without extraction of a physical sample);
- sample preservation and shipment conditions (e.g. storage temperature and/or atmosphere);
- sample mass or volume (and number of increments, if composite samples are required);
- sample containers (e.g. bulk, number of bags or barrels of specified volume and material, etc.);
- notes on the expected degree of heterogeneity of the sampling target and the possible presence of sub-populations;
- final mass or volume of the laboratory sample;
- methods for the physical preparation of the primary sample (in the field or in the lab), e.g.:
    - comminution and/or splitting;
    - subsampling or mass reduction/splitting;
    - drying, sieving, milling, splitting, homogenisation.
- number of test samples, test portions and aliquots to be analysed;
- QA/QC test materials;
- any other relevant information;
- information relating to the safety conditions of the sampling site and specific requirements to consider when handling the samples.

The sampling technician (sampler) should agree with the customer to carry out the sampling in the conditions most representative of routine operations, for example avoiding adverse weather conditions or overlapping with other activities that could affect the effectiveness of the overall procedure (sampling and testing).

The sampler should also ensure that any equipment used for sampling has been decontaminated and verified where possible. Where *in situ* or on-site measuring instruments are required, they will verify that they have been properly calibrated and ensure that the appropriate reference materials or QC test materials are taken for field verification

Once in the field, before proceeding with sampling, the sampler will ensure that the information reported in the agreed sampling procedure is adhered to. If any deviation (s) is (are) made, between the sampling procedure and the actual situation in the field, the sampler, according to his training, will choose between:

- verify that the existing conditions still respect the minimum requirement to perform valid sampling and carry on the activities. In this case the sampler shall report any substantial deviations in the sampling report;

- suspend the activity and proceed with the update of the sampling procedure (after consultation with customer/client and laboratory).

During sampling activities all relevant information shall be described in a Sampling Report. This shall include at minimum:

- sampler identification;
- date, place and time of the measurement and climatic condition if relevant (e.g. in open field sampling);
- reference to the sampling procedure that was implemented;
- conditions of the sampling target: volumes, weights, containers, evidence of potential contamination, etc;
- number and mass/volume of the aliquots taken and their identification (for example by indicating the container number or reporting the sampling scheme applied);
- homogenization and reduction procedure of the primary sample (compared with that specified, see above), if applicable;
- containers and preservatives used;
- any measurement made in the field (*in situ* or on site), and identification of the equipment used;
- any deviations from the previously agreed Sampling Procedure.

The finalized Sampling Report should be approved (signed) by the customer.

It would also be advisable to attach photographic documentation showing at least pictures of the sampling target, subsamples (aliquots) before homogenization, and the laboratory sample.

A Chain of Custody should also be used for sample shipment, reporting shipment condition and time tracking.

4.1.4 Sample handling and transport

In some cases, the laboratory can take the responsibility for providing the suitable sample containers (including unique identification) and any necessary preservatives (including instructions on how to use them).

Persons responsible for handling and transport should be sufficiently trained, including being made aware of any vulnerability to deterioration of the samples / sample packages, and the conditions required during transportation (e.g. temperature). Basically, it could be helpful to support whoever is responsible with specific instructions, especially in cases where the transportation is done by people outside the normal measurement process (e.g. mail or courier services).

4.1.5 Sample reception at the testing laboratory

Under the responsibility of the laboratory manager, the laboratory staff (previously trained on technical guidelines, regulations, recommendations, etc.) are responsible for the following checks and tasks, upon receipt of samples.

- the specific conditions of transport;
- the integrity of the sample containers;
- the accuracy and completeness of the documentation;
- the date and time of sampling should be in agreement with the maximum storage time, where applicable;

- the feasibility of the analytical request (if this is not agreed with the laboratory beforehand).

In some laboratories this responsibility is given to specific persons.

The condition of samples delivered to the laboratory should approximate to that which is defined in technical reference guidelines (for instance technical regulations) where applicable, otherwise samples should be considered as unsuitable for analysis.

It could be useful for the laboratory staff to have a written standard procedure for checking and rejecting samples, including some clear criteria to be followed by the person responsible for judging the suitability for analysis of a sample arriving at the laboratory. It is not always possible to judge the inappropriateness of the sample by its visual inspection upon receipt, and this may require later action based upon quantitative evidence from on-going IMQC, where that is possible (Section 3).

4.1.6 Sample preparation (including physical operations like storage, subsampling and homogenization)

All sample preparation steps should be appropriately designed and conducted, documented and included in the validation of the measurement procedure. Those responsible for sample preparation often include the laboratory staff (often the analysts who carry out the analysis) who are responsible to the laboratory manager. It is the responsibility of the designated laboratory staff to read the relevant requirements of the testing method and/or of the customers before the storage, handling and preparation of the laboratory sample. Such staff should be responsible, competent and reasonably well informed on the relevance of their task. The laboratory manager should draw up the instructions in such a way as to enable the execution of the work effectively and safely.

In some cases, an additional identifiable responsible person in this stage is a representative from an inspection body. The inspector should ensure that the quantity of sample prepared is sufficient for the intended analysis, and for the retention of samples, and that all test samples are derived from the same laboratory sample.

4.1.7 Ultimate responsibility for the overall measurement process

The recommendations set out in the previous paragraphs should be applied, and ideally the laboratory manager should be responsible for the overall measurement process, delegating responsibilities to the relevant people who are involved in the various phases of the process (based on their qualifications and experience). An example of a delegation responsibility framework can be shown in Table 1 (a & b). However, taking responsibility for activities outside the laboratory will often be difficult and requires good communication, not only with customer and samplers, but also with a representative from an inspection body. The laboratory manager should be kept informed, based on clear prior agreements regarding the purpose, sampling procedure and the specific devolved responsibilities, to ensure that nothing can be misunderstood or fall between two non-overlapping areas of responsibilities.

4.2 Organisation of the measurement process

4.2.1 Communication with the customer on the purpose of the measurements

The communication phase between parties (e.g. customer/inspection or regulatory body and laboratory) could be divided into two different steps: a preliminary stage using informal communication (e.g. mobile phones), and a final phase of official and more formal written communication.

The preliminary stage provides a relevant opportunity to discuss / review the design of the sampling procedure, and whether it is deemed satisfactory. The aim is to facilitate the taking, transport and storage of the samples, as well as to inform the laboratory manager of any delay or significant deviation in the execution of the sampling that will affect the subsequent analysis.

The initial communication stage might also identify requirements for additional training, or modifications of the technical procedures within the sampling procedure (including transport and storage) and improved instructions for the use of sampling devices or equipment. This may be particularly useful when the sampling and analyses are carried out by several different organisations.

Every sampling operation requires the drafting of relevant instructions indicating especially the reason for sampling, the place and time of sampling, the definition of the sampling target, the parameters to be measured, the clear labelling of primary samples in accordance with the sampling method, and any special precautions required based upon the information from the customer. Information about the transport conditions and any additional information describing the sampled product, or other information likely to be of assistance to the laboratory in evaluating the results, should also be included. In case of any departure from the recommended sampling procedure, this should be described in detail in the sampling report. A thorough approach, with meticulous attention to detail is essential.

In general, the laboratory should agree with the customer/inspection or regulatory body all operations related to the sampling and/or testing/analysis phase. If the sampling is performed exclusively by the laboratory, then all the information that is relevant for the estimation or verification of the measurement uncertainty is available to the laboratory manager. However, if the sampling is performed by an independent organisation, then this same information should be made available, following previous discussion with the laboratory on the design of the sampling procedure required to enable this.

Given that the sampling technique itself will produce uncertainty in the measurement result, it is essential that the samplers (staff carrying out the sampling) are suitably trained in the procedures used. Moreover, the samplers need to frequently contact the laboratory staff assigned to carry out the analysis, and sometimes also directly with the customer who requested the measurement. This approach is particularly significant in the case of sampling activities performed by several organisations (e.g. governmental or non-governmental agencies, control authorities, quality control laboratories, clients or manufacturers). In such circumstances, precautions should be taken to avoid changes to the sampling procedure which might affect either the analyte concentration in the sampling target or the primary samples, adversely affect the analytical determination, or make the samples inappropriate for the analytical purpose.

Appendix A – Worked Examples

Example A1: Nitrate in glasshouse grown lettuce - Sequential approach to VaMPIS using the Duplicate Method

1. Scope

The validation of a measurement process for the determination of nitrate concentration in glasshouse grown lettuce, using a standard sampling protocol and an analytical procedure (i.e. method) that has previously been validated in isolation. The general approach taken to the validation is that described in Section 2.1 of the current document. The initial part of the validation of this case study, the uncertainty estimation, has already been described in Example A1 of the Eurachem Sampling Uncertainty Guide [3]. Rather than repeating all of that text, a summary is provided here of that aspect, but the reader is referred to the original document for details. This example of VaMPIS has also been reported elsewhere [14].

2. Scenario and sampling target

Nitrate is essential for plant health; however, there are concerns for human health associated with eating food containing elevated levels of nitrate. The concentrations of nitrate in lettuce are regularly monitored in line with EC requirements. Concentration estimates are made for a greenhouse 'bay' of 12,000 to 20,000 lettuce heads, and the result for each bay is 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 [31] (4500 mg kg^{-1}), an estimate of the measurement uncertainty is desirable.

The validation process is described using the following steps from the Sequential VaMPIS flowchart (Figure 2):

Step 1. Specify the measurand of interest in terms of both the analyte and the sampling target (summarised in Table A1.1)

Table A1.1 - Specification of the measurand, which includes the sampling target

Measurand			
Analyte/ Technique	Unit	Sector/ Matrix	Sampling target(s)
Nitrate/hot water extraction and determination by HPLC	mg kg^{-1} as received	Food/ Lettuce	1 bay of iceberg lettuce grown under glass

Step 2: Identify the detailed measurement procedure proposed (including its two components)

Step 2.1 Sampling procedure

The accepted sampling procedure/protocol for this purpose specifies that one composite sample is prepared from 10 heads of lettuce harvested from each bay of lettuce [32]. The lettuces are selected by walking a W shape or five-point die shape through the bay under investigation. This procedure is applied to all bays regardless of their size. In this case the 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.

Step 2.2 Physical sample preparation

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 processor (e.g. Hobart) and macerated to produce a composite sample.

Step 2.3 Analytical method

The analytical method for the determination of nitrate in vegetables and vegetable products [33] had previously been validated using a collaborative trial [34]. In this application, two analytical test portions (10 g) were taken. Each test portion was extracted using hot water and the nitrate concentration determined by HPLC (ultra-violet detector). Quality control samples (spike recovery) were analysed concurrently with the real samples. The original measurement values used for the estimation of uncertainty should have minimal rounding, and should include any values that are less than either zero or the limit of detection.

Step 3. Design the experiment to validate the sampling component of the measurement procedure

The *Duplicate Method* with a full balanced design (Figure 3a) was selected for this validation to limit cost, as it can be implemented by one sampler. It is recognised that this approach, therefore, does not include the contribution to the measurement uncertainty from the between-sampler systematic effects. In many cases it will not be possible to analyse all the primary samples and their duplicates on the same day, as this would, in Example A1 for instance, require 32 analytical measurements. Day-to-day variability could be included in the Duplicate Method design, e.g. if only eight analyses could be performed in one day, then this would be the equivalent of two sampling targets. Other sources of variability could also be included in the design, such as multiple (analytical) operators and multiple instruments. For example, Operator A measures on days 1 and 3, and Operator B measures on days 2 and 4. The analysis of each sample should be randomly assigned on each day, in order to minimise the effects of factors such as instrumental drift.

Given the precautions included in the sample preparation (e.g. freezing within 24 hrs of being taken), it is considered unlikely that physical sample preparation will cause a high level of measurement uncertainty (from either random or systematic sources), and so this component is included in the measurement uncertainty estimate (as sample preparation uncertainty) but does not need to be estimated separately in the validation.

The recommended minimum of eight targets was selected for the uncertainty estimation protocol. These targets were considered typical of bays of greenhouse grown Iceberg lettuce. For each of these bays a second 10-head sample was taken (Sample 2, S2) in addition to the routine sample (Sample 1, S1). This duplicate sample was taken in a way that represented the variation that could occur due to the ambiguities in the sampling protocol, for example positioning of the origin of the W design, and its orientation. The primary and duplicate samples form the second level of the balanced design for each sampling target (Figure A1.1).

It was decided that the inclusion of field blanks would not be applicable to this situation, as the sampling and packing equipment were unlikely to cause either contamination or loss of nitrate from the primary sample.

Step 4. Apply the selected sampling and sample preparation procedures to the sampling target(s)

The sampling procedure was applied by one sampler to all eight bays, and no deviations from the written protocol were reported. The physical preparation of the sample was undertaken by laboratory staff, independently on both duplicated samples for each sampling target.

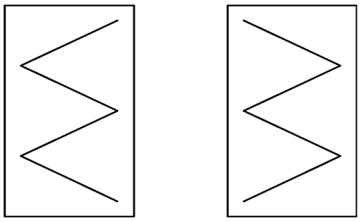

Figure A1.1 - Example of applying the 'Duplicate Method' to the sampling of bays of fresh lettuce. The Duplicate Method can be applied using the W design as a pattern, the protocol stipulates the design but not the position or orientation. The 'W' is equally likely to start on the left or the right. Ten heads are taken along the line of the W to create a composite sample for one target [3]

Step 5. Apply the selected analytical procedure to all of the primary samples taken

Duplicated analyses were made of all the duplicated samples at the lowest level of the balanced design (Figure A1.1). Full analytical quality control (AQC) was applied to the batches of chemical analysis. The resultant measurement values are shown in Table A1.2. No appropriate matrix-matched CRMs were available, so analytical bias was estimated using spike recovery.

Table A1.2 - 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

Step 6. Apply analytical quality control to all of the measurement values in the routine way

The batches of chemical analyses conformed to all the quality control requirements of the laboratory. No significant analytical bias could be detected and so bias correction was not considered necessary for the resultant measurement values, or for the measurement uncertainty estimation.

Step 7. Estimate measurement uncertainty, and its components arising from sampling and analysis (applying ANOVA to the measurement values)

The full details of the measurement values and the subsequent ANOVA (using software RANOVA3) are given elsewhere [3]. For the purpose of this discussion, the estimates of the overall measurement uncertainty and its two components arising from the sampling (including sample preparation) and the chemical analysis (Table A1.3, bottom row) are separated from the overall variability (expressed as total variance and total standard deviation (Total Sdev)).

Robust ANOVA, rather than Classical ANOVA, was employed, because a small proportion (< 10 %) of outlying values was evident in the raw measurement values (Target C, Sample S1 or S2, Table A1.2). Such outlying values would have a disproportionate effect on measurement uncertainty estimated using Classical ANOVA. Robust statistics has proved effective in accommodating such outlying values, at any of the three levels of the experimental design, rather than trying to identify and remove them individually when justified [3]. The robust estimate of the standard deviation (i.e. standard uncertainty) of the measurement procedure overall (s_{meas}) is 360 mg kg^{-1}. The component from the sampling (including sample preparation) (s_{smp}) is 319 mg kg^{-1}, and this contributes around 78 % of the total measurement variance (100 × 22.64 / 28.91, from % of total variance values). In this case the sampling uncertainty is mainly caused by heterogeneity of the analyte (i.e. nitrate) within each sampling target.

Table A1.3 - Output of RANOVA3 [35] software showing robust estimates of measurement uncertainty (Expanded relative uncertainty derived from repeatability) for the case study of nitrate in lettuce (concentration units are mass fraction as mg kg^{-1}). Values are given unrounded for comparison, but will require subsequent rounding

Robust ANOVA

Mean	***4408.3***			
Total Sdev	***670.58***			
	Btn Target	Sampling	Analysis	Measure
Standard deviation	***565.4***	***319.05***	***167.94***	***360.55***
% of total variance	***71.09***	***22.64***	***6.27***	***28.91***
Expanded relative uncertainty (95%)		***14.47***	***7.62***	***16.36***

The repeatability standard deviation values are used as estimates of the measurement uncertainty, either as the standard uncertainty (u_{meas}, estimated as s_{meas}) or the expanded form (for 95 % confidence) relative to concentration (x) using:

$$U' = 100 \times \frac{2\, s_{meas}}{x} \%$$

For this example, the overall measurement uncertainty is estimated as U'_{meas} = 16.36 % = 16 % for routine purposes).

The analytical component of the measurement uncertainty, estimated from repeatability, is U'_{ana} = 7.6 %.

This value is similar to the value of U'_{ana} = 6 % (i.e. u'_{ana} = 3 %), which was previously reported by a separate validation of this analytical procedure [34]. This had been based upon inter-laboratory reproducibility relative standard deviation values [36], in this case of 3.21 % and 2.37 % for two different lettuce types. The U'_{ana} value of 7.6 % estimated here using the Duplicate Method is arithmetically larger than the value from the isolated validation (6 %), which may seem surprising given that the isolated validation value had also included the inter-laboratory component. The increase may be partially caused by analyte heterogeneity within the test material that was used in the Duplicate Method. The isolated approach typically uses a much more homogeneous RM.

Step 8. Judging the fitness-for-purpose of the measurement results by comparing their measurement uncertainty estimates against a target uncertainty

Currently there is no target uncertainty (that includes the sampling uncertainty component) specified by a regulator or a customer, against which this estimated measurement uncertainty (of 16 %) can be compared to judge the fitness for purpose of the measurement results. Various approaches to setting a value for target measurement uncertainty have been discussed [18]. One of these approaches is the Optimised Uncertainty method (Appendix B, [19]), which has previously been recommended [3, Section 16] and applied for this purpose in various sectors [21, 37].

Application of the Optimised Uncertainty method to nitrate in lettuce example

The application of the Optimised Uncertainty method can be made using many software packages, but there is an Excel spreadsheet program (OptiMU) written for this purpose [35]. The input parameters of the Optimised Uncertainty method, and the values used in this example, are shown in Table A1.4.

Table A1.4 - Input data for the Optimised Uncertainty method used to calculate the optimal (target) uncertainty for the example of nitrate in lettuce

Parameter	Units	Value
Sampling Cost	€	40
Analytical Cost	€	40
Consequence Cost	€	5280
u_{smp}	mg kg^{-1}	319.05
u_{ana}	mg kg^{-1}	167.94
u_{meas}	mg kg^{-1}	360.55
Threshold (T)	mg kg^{-1}	4500
Concentration at which to optimise (c_m)	mg kg^{-1}	4871.2

The costs are the commercial unit costs of the sampling and the chemical analysis. When estimating the consequence cost, there are two general options. When a product is erroneously passed as being compliant with the regulations, a false compliance occurs (i.e. a false negative). If this is subsequently discovered, then the producer will be fined and a product recall may be required, and the producer may not retain public support, and so experience a drop in sales or a sharp deduction in share price. Past examples can be utilised when evaluating this parameter. Alternatively, when a product is wrongly classed as being non-compliant with a regulatory threshold, a false non-compliance scenario occurs (i.e. a false positive). This typically results in the unnecessary rejection of a batch of product. The cost here is typically evaluated as the cost of the batch. In this example, the consequence cost (i.e. € 5280) is calculated for a *false non-compliance* (i.e. false positive) decision based upon the value of an entire batch of 12,000 heads of lettuce at € 0.44 (all prices applied at time of validation in 2004).

The uncertainty values are derived from the robust ANOVA output, discussed above. The threshold concentration (T) is that specified in EU regulations [31].

The concentration at which the system is to be optimised (c_m) is generally selected so that there is an appreciable probability of misclassification. Previous applications of the Optimised Uncertainty method have utilised a range of criteria for the setting of c_m (e.g. $1.1T$ = 4,950

mg kg^{-1}) [38]. For this investigation, the level of c_m was set at a hypothetical enforcement limit of nitrate in lettuce [23]. The relative expanded analytical uncertainty (U'_{ana}) was already estimated to be 7.62 %. The minimum concentration which would indicate that the nitrate concentration was greater than the threshold ($c_m - U_{ana} = T$) was calculated to be 4,871 mg kg^{-1} ($T + U_{ana}$ at c_m, = 4500 + (0.0762 × 4871). Interestingly, this value is similar to two other values that can be calculated for setting c_m using alternative approaches, which are the median value of non-compliant measurements (4,891 mg kg^{-1}) and the value of 1.1T (4,950 mg kg^{-1}).

Step 9. Assess the extent to which fitness for purpose has been achieved (comparing the experimental measurement uncertainty against the optimal target uncertainty value)

The results of this application of Optimised Uncertainty method (Figure A1.2) shows the example-specific version of the general case (Figure B1.1), where the total cost is expressed formally as 'expectation of loss' E(L). The Actual measurement uncertainty (as s_{meas} = 361 mg kg^{-1}, E(L) = € 873) is clearly well above the optimal (used as target) measurement uncertainty (as s_{meas} = 184 mg kg^{-1}, E(L) = € 395). In units of relative expanded uncertainty, the U'_{meas} values of these two points are 16.4 % and 8.3 % respectively. This indicates this measurement procedure produces measurement values that are *not* fit for purpose, and is therefore not eligible for validation in its current form.

However, Figure A1.2 also suggests that if the measurement uncertainty could be reduced from 361 mg kg^{-1} to around 184 mg kg^{-1} then fitness for purpose could be achieved at Target measurement uncertainty, and thereby validate this measurement procedure. The calculation also indicates that an overall saving of € 478 (€ 873 − € 395) per batch would then be made by reducing the risk of a false non-compliance classification.

Step 10. Modifying the measurement procedure to achieve fitness for purpose (if required)

If a measurement procedure is shown *not* to be fit for purpose, due to the measurement uncertainty being substantially higher (or lower) than the target uncertainty (however set), it is possible to calculate how the procedure can be modified to produce a value for measurement uncertainty that will be fit for purpose. The second part of the Optimised Uncertainty method can be used to calculate whether it is more cost-effective to modify either the sampling or the analytical procedure. For this example, applying four Equations (B1.5 to B1.8 in Appendix B) using the respective cost of sampling and analytical procedures. The Optimised Uncertainty method calculates that the most cost-effective way to reach the target measurement uncertainty overall, is to reduce the sampling uncertainty by a factor of 2.14 (from 319 to 149 mg kg^{-1}). This uses a model from Sampling Theory that predicts that the square of fundamental sampling error (and hence usually s_{smp}) is inversely proportional to the sample mass [3, p24]. The Optimised Uncertainty approach also predicts that this reduction in sampling uncertainty could probably be achieved by increasing the expenditure on sampling by a factor of 4.57 from € 40 to € 183 per batch. In practical terms, an increase in the sample mass by a factor of 4 (i.e. taking a 40-head composite sample in place of a 10-head composite) is predicted to reduce the sampling uncertainty by a factor of 2 (i.e. √4). Given that sampling uncertainty contributes around 78 % of the measurement uncertainty, then this modification of the procedure should then also reduce the overall measurement uncertainty by a similar factor.

An experiment was made to investigate the actual reduction in sampling uncertainty (and hence measurement uncertainty) that would be achieved by increasing the number of lettuce heads within the composite sample from 10 to 40 [23]. The consequent reduction in sampling uncertainty was found to be by a factor of 1.8, which is similar but arithmetically slightly lower than the factor of 2 predicted by sampling theory. When this factor of 1.8 is applied to the example under discussion, this reduces the sampling uncertainty as s_{smp}, from 319 to 177 mg kg^{-1}. When recombined with the analytical component, this gives a reduction of overall measurement uncertainty (as s_{meas}) from 360 to 244 mg kg^{-1} (yellow spot on Figure A1.2). As expanded relative uncertainty, this is a reduction of U'_{meas} from 16.4 % to 11.1 %. This modified measurement uncertainty is close to (33 % above) the *Optimal* (i.e. target) measurement uncertainty of 184 mg kg^{-1}, and gives a reduction in the cost (i.e. expectation of loss) of around

€ 375 per target (from around € 873 to € 498), which is close to (78 % of) the optimal saving of € 478. The measurement procedure can now, therefore, be said to be validated, as its measurement results are effectively fit for purpose.

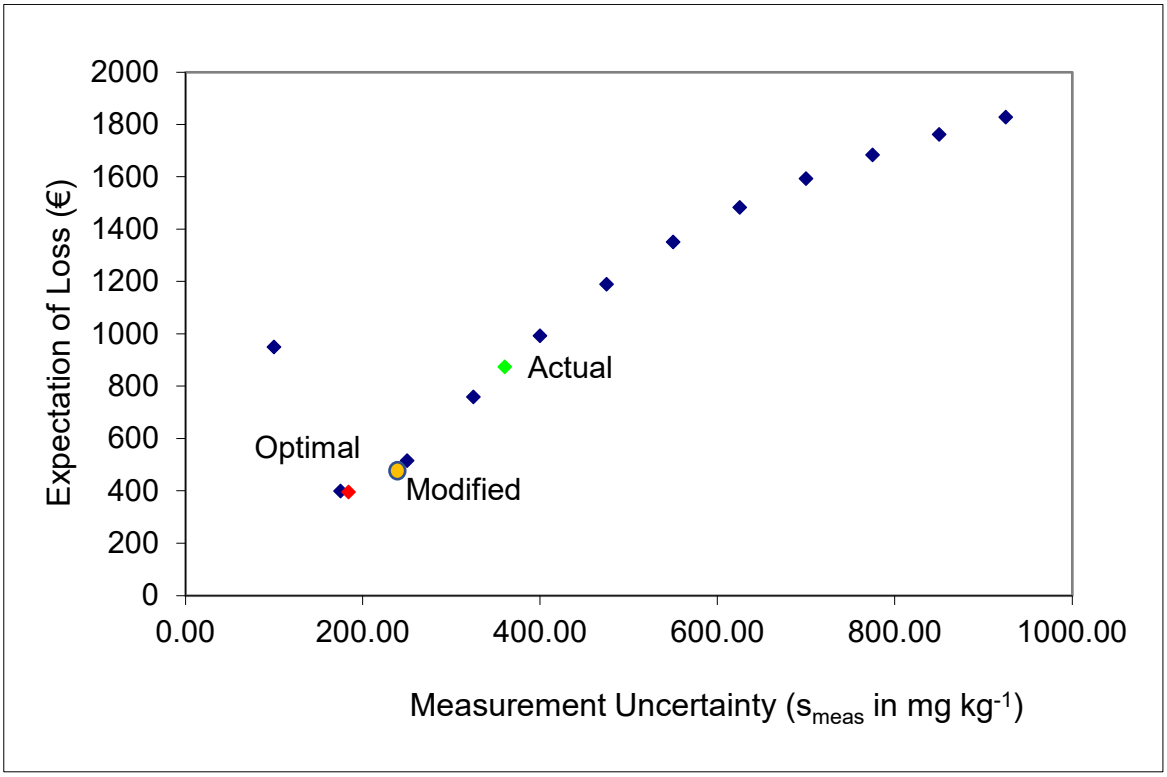

Figure A1.2 - The Optimised Uncertainty curve for the example of Nitrate in Lettuce, showing the estimate of the Actual measurement uncertainty (as s_{meas} = 361 mg kg^{-1}, E(L) = € 873, shown as green spot) is far above the Optimal (used as Target) measurement uncertainty (as s_{meas} = 184 mg kg^{-1}, E(L) = € 395, as red spot). This indicates that the current measurement procedure is *not* fit for purpose, and cannot be validated in its current form. A Modified measurement uncertainty (as s_{meas} = 244 mg kg^{-1}, E(L) = € 500, as yellow spot) is discussed under Step 10 [14]

Step 11. Review fitness for purpose of analytical procedure

The Optimised Uncertainty method suggests that the analytical procedure that was validated previously (with U'_{ana} of ~ 6 %) achieved an overall measurement uncertainty that is close enough to the optimal measurement uncertainty for the current validation of the overall measurement system. Additionally, the Optimised Uncertainty calculations (Equations B1.6 & B1.8, Appendix B) also suggest that a slight decrease in the analytical measurement uncertainty (by a factor of 1.56) from 168 to 108 mg kg^{-1} (i.e. U'_{ana} from 7.6 % to 4.9 %) would also be beneficial to achieving optimal (i.e. target) uncertainty. It estimates that this reduction in U'_{ana} could be achieved by increasing the analytical expenditure by a factor of 2.4 from € 40 to € 96. This expenditure could be targeted to lower U'_{ana} by reviewing and improving some of the seven performance characteristics already discussed (Section 2.2). Interestingly, if this lower U'_{ana} was achieved by improving the analytical method, the overall measurement uncertainty would be predicted to drop further to 207 mg kg^{-1} (U'_{meas} = 9.4 %), which is even closer to the Optimised Uncertainty (used as U'_{TU}) of 184 mg kg^{-1} (U'_{meas} = 8.4 %). Target measurement uncertainty is defined as a maximum value, but in this case an actual measurement uncertainty slightly over the value has a similar financial effect to it being slightly below. Because the analysis only contributes around 22 % of the measurement uncertainty, its marginal reduction would have much less effect than the 5-fold increase in sampling expenditure already

implemented. Overall, therefore, the current analytical procedure can be considered suitable for inclusion in this measurement procedure.

Summary

The process for the sequential validation of a measurement procedure (Section 2.1 of main text) has been applied to the determination of nitrate in lettuce. The validity of the proposed measurement procedure was judged by the uncertainty of the resultant measurement values, and found to be *not* fit for purpose, as the measurement uncertainty of 16.4 % was around twice the target uncertainty of 8.3 %. The target uncertainty was not specified by the regulator or the customer, so it was calculated using the Optimized Uncertainty method. The Duplicate Method, that was used to estimate the actual measurement uncertainty, also showed that the dominant source of the measurement uncertainty was the sampling (U'_{smp} of 14.5 % contributed 78 % to the overall measurement uncertainty). Sampling theory was used to predict that increasing the number of increments in each composite sample from the originally specified 10 up to 40, would reduce the overall measurement uncertainty down to 11.1 %. This modification was considered sufficient to make the measurement values fit for purpose, and for the measurement procedure to be validated. This validation does not include the contribution to measurement uncertainty from the between-sampler bias that could be provided by a Collaborative Trial in Sampling (CTS). However, given the dominant contribution to measurement uncertainty from the sampling caused mainly by the heterogeneity of the nitrate, this single-sampler approach may give a sufficiently realistic estimate of measurement uncertainty for the purpose of validation.

Applicability of the validation

This validation would be expected to be generally applicable to glasshouse grown batches of lettuces of any variety. This is largely because lettuce targets are all grown in such a way as to meet the same specified threshold and would therefore tend to be generally similar in composition. The validity of this assumption should be tested by ongoing IMQC results over batches grown over a range of different conditions. Applying this measurement protocol to other food crops would require at least verification and possibly full validation, due to the likely different characteristic of the spatial distribution of the analyte (nitrate) in different crops.

Example A2: *In situ* measurement of total lead in top soil - Simultaneous approach to VaMPIS using the Duplicate Method

1. Scope

The validation of an *in situ* measurement process for the determination of total lead (Pb) concentration in top soil, using an integrated measurement procedure (i.e. both sampling and analytical procedures). The general approach taken is therefore that of simultaneous validation (Section 2.3.3). The validation described is largely based upon an estimate of the measurement uncertainty that is typical of a routine operation [39]. Fitness for purpose of *in situ* measurements is evaluated where these are made for two possible purposes: (1) the geochemical mapping of Pb concentration in top soils across the test site (several sampling targets), and (2) the classification of the land for Pb concentration that possibly exceeds a regulatory threshold value.

2. Scenario and sampling target

Lead is a heavy metal that is very toxic to human, animal and plant life, and therefore its concentration is regulated in soils in many countries. The case study was undertaken at a site of approximately two hectares in Wirksworth, Derbyshire, UK [27], that was suspected as formerly being the location of a medieval Pb smelter. The site was therefore expected to have soils with a wide range of different, but still elevated, Pb concentration.

The validation process is explained following the steps described in the Sequential VaMPIS flowchart (Figure 2), and in Section 2.3.3.

Step 1. Specify the measurand in terms of both the analyte and the sampling target: this is summarised in Table A2.1

Table A2.1 - Specification of the measurand, which includes the sampling target

Measurand			
Analyte/ Technique	Unit	Sector/ Matrix	Sampling target(s)
Total Pb by hand-held Portable XRF applied *in situ*	mg Pb per kg of dried soil. (i.e. mg kg^{-1})	Top soil (nominal depth 0 - 150 mm)	30 x 30 m areas of soil (within a site with 24 such targets)

Step 2: Identify the detailed measurement procedure proposed, including its two main components of sampling and analysis

Step 2.1 Location of sampling targets

The irregular grassy site, used for grazing animals, was covered by a regular sampling grid of 24 sampling targets with a 30 m spacing (Figure A2.1). The location of each *in situ* measurement was at the centre of a square of side 30m, located using measuring tape with an estimated spatial uncertainty of ± 2 m.

Step 2.2 Instrumental procedure

At each measurement location, prior to measurement, an area of turf 150 x 150 mm by 30 mm deep was removed using a spade to reveal bare soil (i.e. not part of the sample). The material removed, which was not part of the sample, was set aside and replaced after the completion of measurements. The hand-held portable X-ray fluorescence spectrometer (pXRF) was then placed manually at the centre of the bare earth, gently in contact with the soil. A 75 second

measurement cycle was performed using the instrumental calibration designed for soils by the manufacturer of the pXRF, with its analytical performance and instrumental settings described elsewhere, both for laboratory measurements, addressing *analytical sensitivity* and *selectivity* [40] and also for *in situ* use in the field [27]. A particular issue for *selectivity* of Pb determination by XRF is the potential interference on the Pb-Lα x-ray line from high concentrations of arsenic (As). The effectiveness of the correction of this interference can be assessed for this site by estimating *trueness* as *bias* using a CRM with high concentrations of both As and Pb (e.g. NIST 2710). The *limit of detection* for Pb was estimated to be 39 mg kg^{-1} [27], which is well below the minimum concentration measured by pXRF at this site (~330 mg kg^{-1}). The upper limit of the *working range* is well above the corresponding maximum measured concentration (~11,000 mg kg^{-1}).

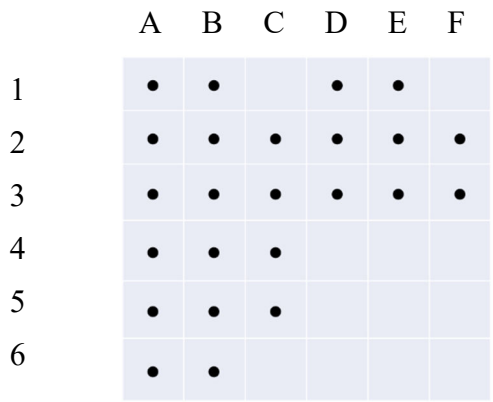

Figure A2.1 - Map of the grid of 24 square sampling targets each measuring 30 m x 30 m, with a central measurement location, laid out across the test site (within its irregular site boundary)

Step 3. Design the experiment to validate the measurement procedure (sampling and analysis)

The overall measurement uncertainty was estimated by applying the Duplicate Method with a simplified balanced design (Figure 3c), with single measurements made on duplicate 'samples' at all 24 sampling targets. This reduced the measurement time at each sampling target, and thereby enabled the procedure to be applied to more targets. The minimum number of duplicates required is 8 [3], but the increased number of 24 duplicates in this case gives much smaller confidence intervals on the estimated measurement uncertainty values [8].

A first sampling location was located as being at the centre of the sampling target using the written measurement procedure, and an instrumental measurement (i.e. chemical analysis) taken. A second (duplicate) location was then located 2 m away from the first location, to represent an equally likely reinterpretation of the measurement procedure (made manually using measuring tape and canes). This spatial uncertainty is similar to the 5m typical of standard GPS [41], but use of differential GPS can reduce this value to a few cm. In that case, the spacing of the duplicate sample (i.e. measurement) can be reduced to broadly reflect that situation. However, small differences in the separation of duplicate samples does not usually result in significant changes in the estimate of sampling uncertainty [42].

Steps 4 & 5. Apply the selected measurement procedures (sampling and analysis) to the sampling target(s)

Because of the integrated nature of *in situ* measurements, all of the steps were performed by one operator on all 24 of the sampling targets, with no reported deviations from the written procedure. No significant instrumental drift was detected by measurement of matrix-matched RMs before and after the survey. This supported the implementation of a sequential rather than

a random measurement order of each sampling target across the site. The resultant measurement values are shown in Table A2.2.

One component of analytical bias was estimated by measuring pressed pellets made from a series of 4 matched soil RMs (3 in-house RMs and one CRM NBS/NIST 2710), both immediately before and just after the field measurements, to also estimate any potential instrumental drift.

Table A2.2 - *In situ* measured Pb concentration values [43]

Sample I.D.	S1Pb (mg kg^{-1})	S2Pb (mg kg^{-1})
A6	1005	1633
A5	4631	3723
A4	1415	2264
A3	865	1350
A2	2899	2216
A1	721	1758
B6	2122	1014
B5	1321	1043
B4	3348	3904
B3	11543	5570
B2	2904	2833
B1	2617	2762
C5	976	786
C4	6127	3874
C3	331	576
C2	12878	8948
D3	3246	4332
D2	9006	6098
D1	1936	1989
E3	5811	6289
E2	4611	2880
E1	1326	1442
F3	1215	2713
F2	2070	2305

A more comprehensive estimate of measurement bias was made by also taking physical samples of the soil at the same locations, after the *in situ* measurements were made. Core samples of diameter 25 mm and depth 0 – 150 mm were taken using a soil auger, and the resultant primary samples dried overnight at 65 C, disaggregated, sieved to retain the size fraction less than 2 mm (by which soil is defined [44]) and then ground to < 100 µm in an agate ring mill. Duplicate test portions of these test samples were then analysed for Pb in a full balanced design, using ICP-AES after acid digestion (nitric and perchloric), with full AQC including analytical

duplicates and the 6 CRMs (BCR 141, BCR 142, BCR 143, NBS2709, NBS2710, NBS2711) to enable the estimation of analytical bias, and also to give traceability. Similar test portions were also taken for direct measurements by lab-based pXRF, also from this full balanced design, mainly for the purpose of estimating the purely analytical contribution to the measurement uncertainty [43].

Step 6. Apply analytical quality control to all of the measurement values in the routine way

Ten field measurements of NBS/NIST 2710 had a mean value of 4935 mg kg^{-1}, which against a certified value of 5532 mg kg^{-1}, gives a statistically significant estimated bias ($\pm$ s/$\sqrt{n}$) of -10.8 ± 0.7 % [45]. The equivalent value for significant bias of pXRF in the lab (*ex situ*) was -11.5 ± 1.1 %, and for the ICP-AES *ex situ* procedure was -5.5 ± 0.05 % using the 6 CRMs. The negative bias for Pb indicates that the possible spectral interference (e.g. from As) has been effectively corrected and is not affecting the selectivity. The analytical duplicates were included in the full balanced design and processed using ANOVA, interpreted below.

Step 7. Estimate measurement uncertainty, and its components arising from sampling and analysis, by applying ANOVA to the measurement values

Because a simplified experimental design was used (Figure A2.2), the measurement values (Table A2.2) were entered in a particular version of the appropriate software RANOVA3 [35].

Table A2.3 - Output of RANOVA3 software [35] showing classical and robust estimates of measurement uncertainty (Expanded relative uncertainty derived from repeatability) for the case study of *in situ* measurement of lead in top soil (concentration units are mass fraction as mg kg^{-1})

Classical ANOVA

Mean	***3275.5***	No. Targets	***24***
Total Sdev	***2797.3***		
	Btn Target		Measure
Standard deviation	***2494.8***		***1265.1***
% of total variance	***79.55***		***20.45***
Expanded relative uncertainty (95%)			***77.25***
Uncertainty Factor (95%)			***1.8514***

Robust ANOVA

Mean	***2856.6***	
Total Sdev	***2050***	
	Btn Target	Measure
Standard deviation	***1893.5***	***785.61***
% of total variance	***85.31***	***14.69***
Expanded relative uncertainty (95%)		***55.00***

The robust estimate of the overall measurement uncertainty (measurement uncertainty as the relative expanded uncertainty *U'*) is given as 55 % in the right-hand column in the lower part of Table A2.3. This robust estimate is more applicable than the classical estimate (77.3 %),

because the frequency distribution of the measurement values is positively skewed, and clearly not Gaussian (Figure A2.2a). It is not possible initially to separate the two components arising from sampling and chemical analysis using this experimental design. However, the 'sampling' component of measurement uncertainty (estimated as U_{smp}) can be separated retrospectively by subtracting an external estimation of the analytical component from the measurement component using their variances (U'^2_{ana} *and* U'^2_{meas} respectively in Equation A2.1).

$$U'_{smp} = \sqrt{U'^2_{meas} - U'^2_{ana}} \qquad \text{Equation A2.1}$$

An external value of U'_{ana} as 3 % was estimated using the additional *ex situ* pXRF measurements made in the laboratory on the prepared versions of removed primary samples from the same 24 targets, in fully balanced experimental design (i.e. with duplicated analyses on 10 duplicated samples) [27]. This approach assumes that the instrumental performance of pXRF is similar when used *in situ* and *ex situ.*

Substituting these values into Equation A2.1, applied to relative expanded measurement uncertainty values gives:

$$U'_{smp,in\ situ} = 54.9\ \% = (\sqrt{55^2 - 3^2})$$

This clearly shows that the instrumental analysis contributes a negligible proportion (<0.3 %, 100 x $3^2/55^2$) to the overall variance from the measurement uncertainty.

Because the measurement uncertainty value is large (u' > 15 – 20 %) and the frequency distribution is positively skewed (Figure A2.2a), measurement uncertainty can be expressed more accurately as the **Uncertainty Factor** ${}^{F}U$ [46]. Robust ANOVA is adjusted to accommodate up to 10 % of outlying values, but this will not be as effective for the high level of skew shown in Figure A2.2a. The value of ${}^{F}U$ can be calculated from the standard deviation of the $\log_e$-transformed measurement values ($s_{G,meas}$), shown in Figure A2.2b, using:

$${}^{F}U = \exp(2s_{G,meas}) \qquad \text{Equation A2.2}$$

This $\log_e$ transformation and calculation is performed automatically by RANOVA3, and the value of ${}^{F}U$ for this study is 1.85 (Table A2.3), and $s_{G.meas}$ of 0.308.

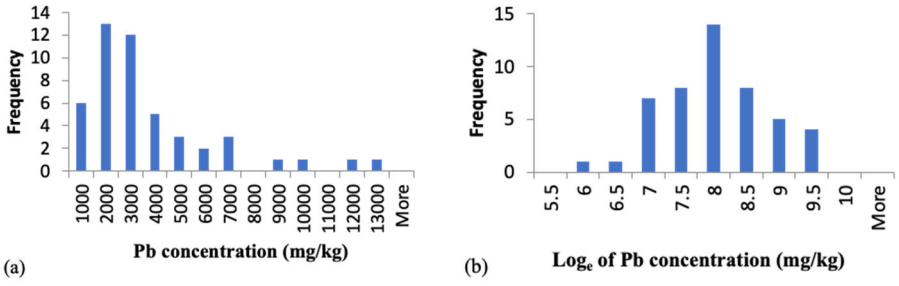

Figure A2.2 - The frequency distribution of the 48 measurement values in Table A2.2 showing (a) a positively skewed probably log-normal distribution, which is confirmed by (b) an approximately normal distribution of the natural logarithms of the values

The Duplicate Method estimates only the random components of the measurement uncertainty, so consideration has also to be given to the systematic components, such as measurement bias.

Inclusion of systematic effects within measurement uncertainty estimates

The measurements made on NIST 2710 gave an estimated analytical bias of −10.8 % (±0.7%). However, the physical form of CRMs is very different from those of the test materials measured *in situ*, as discussed below.

It is therefore more realistic to estimate the bias by comparison of the *in situ* measurement values against the *ex situ* measurement values made on physical samples taken at the same locations/points, as shown in Figure A2.1, and these are shown in Table A2.4.

Table A2.4 - Measurements of Pb concentration in soil made *in situ* using pXRF (with their standard uncertainty, *u'* of 27.5 %) compared against those made *ex situ* on extracted samples using ICP-AES (*u'* of 35.8 %)

Target ID	Ex situ ICP-AES	*u'* ICP (35.8 %)	In situ P-XRF	*u'* pXRF (27.5 %)
A6	7340	2628	1319	363
A5	8815	3156	4177	1149
A4	1522	545	1840	506
A3	1290	462	1108	305
A2	9340	3344	2547	700
A1	3080	1103	1240	341
B6	4180	1496	1568	431
B5	1926	690	1183	325
B4	3670	1314	3626	997
B3	6718	2405	8555	2353
B2	5630	2016	2869	789
B1	3630	1300	2690	740
C5	6880	2463	881	242
C4	9370	3354	5002	1376
C3	1522	545	454	125
C2	21877	7832	10919	3003
D3	5230	1872	3788	1042
D2	18784	6725	7556	2078
D1	2800	1002	1963	540
E3	10584	3789	6050	1664
E2	7316	2619	3745	1030
E1	2235	800	1384	381
F3	3860	1382	1964	540
F2	5210	1865	2188	602

The relationship between the two sets of Pb measurement values (Figure A 2.3) shows that the *in situ* values are generally much lower than the *ex situ* values. The relationship was modelled using a maximum likelihood method of fitting a linear function when there is uncertainty on both the independent and dependent variables, known as FREML, that allows for the individual measurement uncertainty of both sets of measurement values [47]. The relative standard uncertainty (*u'*) for the *in situ* values is 27.5 % (i.e. = *U'* of 55 % divided by 2) and for the *ex situ* measurements is 35.8 %, which is 71.5% / 2 (where 71.5 % is the value of *U'* derived from ANOVA of the full balanced design on 10 duplicated physical samples [43]).

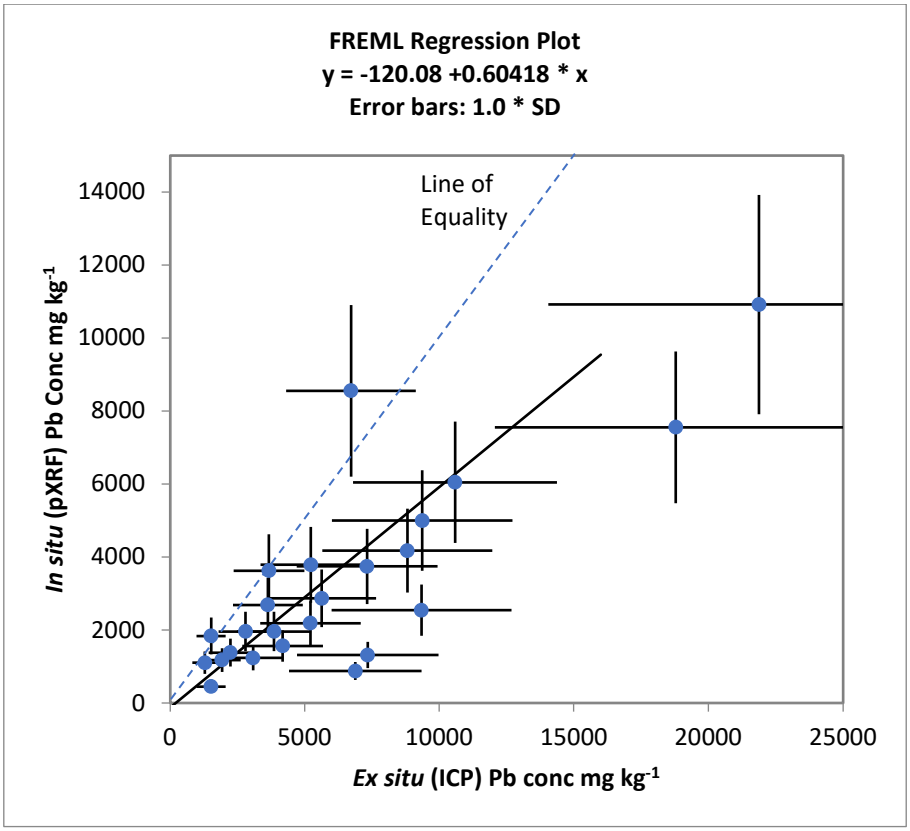

Figure A2.3 - The relationship between *in situ* and *ex situ* measurement values for Pb concentration, modelled using FREML, showing a general negative bias for the *in situ* measurement, estimated by the statistical model as a rotational bias of – 50 %

The general equation describing the statistical model of the bias between the two sets of Pb concentration [Pb] measurements is:

$$[Pb]_{in\ situ} = \beta \times [Pb]_{ex\ situ} + \alpha \qquad \text{Equation A2.3}$$

The slope coefficient of linear model β gives the rotational component of bias and the intercept coefficient α translational component, as described in Figure 5 in Section 2.4.3.

For this case study, the modelled equation of the relationship is:

$$[Pb]_{in\ situ} = 0.60\ (\pm 0.09) \times [Pb]_{ex\ situ} - 120\ (\pm 288)$$

Both of the model coefficients have a standard error (se_β and se_α, shown in parentheses) which allows their statistical significance to be assessed. The standard error of the intercept coefficient is over twice as large as the coefficient itself ($t = 288/120 = 2.4 > t_{0.05(2),22(\text{i.e. n-2})} = 2.074$)

meaning that the intercept coefficient is not statistically significant (i.e. not different from zero) and that there is therefore no detectable translational bias.

The slope coefficient is statistically significant (t = 0.09/0.60 = 0.15 < $t_{0.05(2),22}$ = 2.074), so the rotational bias can be calculated as -40 % (± 9 %) (i.e. (1 – 0.6) x 100).

Possible causes of this large measurement bias have been identified as: a) soil moisture; b) material/particles > 2 mm diameter; c) surface roughness in the pXRF 'undisturbed sample'; d) differences in depth between the undisturbed virtual sample for *in situ* pXRF (~1 mm) and the removed *ex situ* field sample for ICP-AES (150 mm) [27]. A profile of Pb concentration of the top 62 mm depth of two cores, using *in situ* pXRF on eleven 6 mm slices of the core, showed no detectable systematic change of concentration with depth [27]. This finding suggests that the discrepancy of depth between the *in situ* and *ex situ* measurements might not be a dominant cause of systematic error between them.

Treatment of systematic component of measurement uncertainty for *in situ* measurements

There are two options for how to treat the estimate of systematic effects, such as analytical bias in the estimation of measurement uncertainty, but not yet a consensus on the recommended option.

The first option is to 'correct' *in situ* measurements ($[Pb]_{pXRF,corr}$) to agree with *ex situ* values by applying a rearrangement of the bias model. This assumes that the measurand in this case is defined in terms of the analyte concentration reported on a dry-weight basis (as stated in Table A2.1). For this case study the correction equation for the measurement bias (rearranging Equation A2.3) becomes:

$$[Pb]_{pXRF,corr} = \frac{[Pb]_{pXRF,raw} - \alpha}{\beta}$$

In this case, inserting the value of the slope coefficient, and omitting the non-significant intercept coefficient gives:

$$[Pb]_{pXRF,corr} = \frac{[Pb]_{pXRF,raw}}{0.60}$$

The uncertainty of this correction (se'_β = 0.09) can be combined (as a relative percentage 9 %), into the measurement uncertainty value of U' = 55 % (u' = 27.5 %) by the approach described in [3, Section A2/6.4], all expressed as relative standard uncertainty:

$$u'_{corr} = \sqrt{(u'^2 + (se_\beta^2)}$$

$$u'_{corr} = \sqrt{(27.5^2 + 9^2)} = 28.9\ \%$$

$$U'_{corr} = 57.9\ \%$$

Another alternative is to add the uncertainty from the correction into the measurement uncertainty when expressed as the ${}^{F}u$, as $s_{G,meas}$ (= ln(${}^{F}u$)), using an approximation which is applicable when s'_{bias} is less than 0.2, which in this case it is (se'_β = 0.09) [48].

$$S_{G,meas,corr} = \sqrt{s_{G,meas}^2 + \left(s'_{bias}\right)^2} \qquad \text{Equation A2.4}$$

$$S_{G,meas,corr} = \sqrt{0.308^2 + 0.09^2} = 0.321$$

The expanded uncertainty factor ${}^{F}U$, previously calculated for this study as 1.85, can then be recalculated using Equation A2.2, to give a slightly increased value of 1.90 (= exp (2 x 0.321)).

The second option for inclusion of systematic effects into the measurement uncertainty estimate is not to correct the measurement values for the bias, but to add the overall bias, and its uncertainty, into the measurement uncertainty ([3, Section A2/6.4 page 50], again all expressed in terms of relative percentage. For this case study, this would give the large U' value of 98 %, which would not be realistic for the interpretation of results in this case.

$$u'_{corr} = \sqrt{(u'^2 + ((\beta - 1)^2) + (se_\beta^2)}$$

$$u'_{corr} = \sqrt{(27.5^2 + 40^2 + 9^2} = 49\%$$

$$U'_{corr} = 98\ \%$$

Step 8. Judging the fitness-for-purpose of the measurement results by comparing their measurement uncertainty estimates against a target uncertainty.

No value has been set externally for the target uncertainty of *in situ* measurements of Pb in soil. However, there are two possible ways of setting the target uncertainty, which depend on the stated purpose of making the *in situ* measurements.

Purpose 1: Mapping of the analyte concentration across the site to identify the areas of higher and lower concentration of the contaminant

The fitness for purpose criteria for geochemical mapping is that the measurement uncertainty (arising from both sampling and analysis) should not exceed 20 % of the total variance of the analyte concentration across the area to be mapped, which includes the variance between the different sampling targets [49 and 3, Section 16.2]. Using 20 % of total variance, the target standard uncertainty ($u_{meas.Target}$) is given by:

$$u_{meas.Target} = \sqrt{0.2 \times s_{Total}^2}$$

This robust estimate of total standard deviation for the site in this case study was 2050 mg kg^{-1} (Table A2.3). The target measurement uncertainty, expressed as $u_{meas.Target}$, would therefore be 917 mg kg^{-1}. In terms of expanded measurement uncertainty this could be expressed as either U = 1834 mg kg^{-1}, or U' = 64 %.

8.2. Purpose 2: Classification of the land against a threshold limit of contamination

The calculation of the optimal level of measurement uncertainty for both *in situ* and *ex situ* measurements for systems similar to that used in this case study have been described in detail [50]. That optimal value can then be used as the target uncertainty. The parameters for the Optimized Uncertainty calculations (Appendix B), taken from the ANOVA without correction (Table A2.3) are shown in Table A2.5.

Table A2.5 - Input data for the calculation of Optimal Uncertainty, hence target measurement uncertainty. (*the value of μ_{meas} that includes the correction of bias (Equation A2.4) would be 827.0 mg kg^{-1})

Item	Value	Units
Sampling cost (each)	29	€
Analytical cost (each)	12	€
Consequent cost per location, for false positive classification (i.e. unnecessary remediation)	10000	€
u_{smp}	784	mg kg^{-1}
u_{ana}	43	mg kg^{-1}
u_{meas} *	786	mg kg^{-1}
Threshold value of concentration	2000	mg kg^{-1}
Concentration at which to optimise	2200	mg kg^{-1}

The threshold value for the maximum concentration of Pb in soil in this particular situation of green open spaces at the time of the experiment (1995) was 2000 mg kg^{-1} [51]. The costs of the sampling and the instrumental analysis are subdivided in this calculation and are based upon the total cost of labour and equipment, apportioned to each measurement (Table A2.5). The consequence costs are based on the effect of a false positive classification, for this example. This occurs when the measured concentration value is above the threshold value, but the true value (i.e. value of the measurand) is below the threshold value. The cost is calculated from the volume of soil in the falsely classified sampling target that is unnecessarily remediated, at the prices then prevailing [50].

The value of optimal measurement uncertainty, used as target uncertainty, can be calculated using the Equation B1.1, the derivation of which is described in Appendix B.

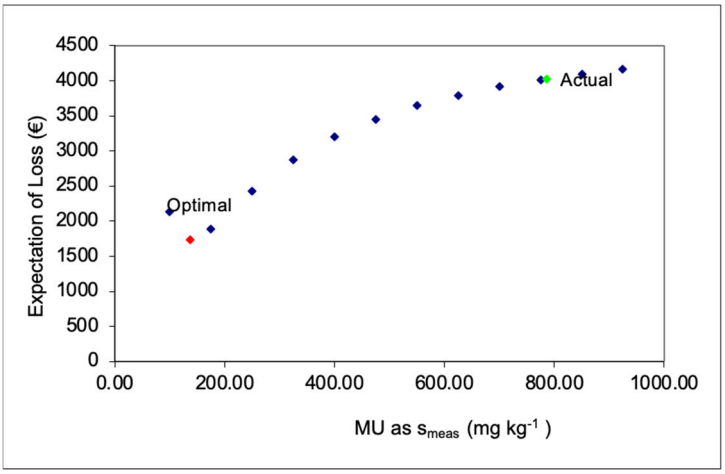

Figure A2.4 - Relationship between the overall cost (expressed as Expectation of Loss from Equation B1.1) and the measurement uncertainty (expressed as $s_{meas} = u_{meas}$) showing that the actual measurement uncertainty (786 mg kg^{-1}, ●) is far higher than optimal measurement uncertainty (138 mg kg^{-1}, ●) required to achieve fitness for purpose for *in situ* measurements for the purpose of land classification

Step 9. Assess the extent to which fitness for purpose has been achieved by comparing the experimental measurement uncertainty against the optimal (i.e. target) measurement uncertainty value.

Purpose 1 Geochemical Mapping

Using the Target u_{meas} of 917 mg kg^{-1} (based upon fitness for purpose criterion of 20 % of total variance), the actual robust estimate of u_{meas} at 786 mg kg^{-1} (Table A2.3) indicates that both the measurement results, and therefore the measurement procedure, *are* fit for that purpose. In terms of the relative uncertainty, the same conclusion of being fit for purpose applies as the actual measurement uncertainty of 55 % (Table A2.3) is less than the Optimal/Target measurement uncertainty of 64 % (using robust estimates throughout). This assumes that the Target measurement uncertainty is more of a preferred value than a rigorous maximum target. The fact that the actual measurement uncertainty is below the Target measurement uncertainty is not a deficiency but actually beneficial as it further improves the reliability of geochemical mapping.

Purpose 2: Classification of land against a threshold

The relationship between Cost (Expectation of Loss, or EoL) and measurement uncertainty (Figure A2.4), shows that the actual measurement uncertainty of (u = 786 mg kg^{-1}, U' = 55 %) is far above the optimal value (u = 138 mg/kg, U' = 9.7 %). The equivalent cost for the actual measurement uncertainty (EoL = € 4026) is more than twice as high as that at the optimal value of measurement uncertainty (EoL = € 1739). This means that the *in situ* procedure is currently *not* fit for the purpose of classifying the Pb concentration in the soil against threshold (T) of 2000 mg kg^{-1} at this site. The achievement of fitness for purpose would not be improved if allowance for the effect of bias correction (Equation A2.4) was applied to calculate the actual measurement uncertainty to give the quite similar value of 827 mg kg^{-1}**.**The more conventional approach would be to use *ex situ* measurements based upon extracted primary samples for this purpose. However, the measurement uncertainty of the *ex situ* approach applied here with analysis by ICP-AES, gives an even larger measurement uncertainty of 71.5 %, again dominated by the sampling uncertainty (99.7 % of measurement uncertainty) caused predominantly by the heterogeneity of the Pb concentration within these targets. This *ex situ* measurement option can also be shown, by the same procedure, to be far above the target measurement uncertainty that is required to achieve fitness for purpose.

Step 10. Modifying the measurement procedure to achieve fitness for purpose, if required

Purpose 1: The *in situ* procedure is already fit for this purpose of geochemical mapping, so nominally does not require any modification. This is evident from the geochemical map that was prepared for this site (Figure A2.5a), which shows that the pXRF measurement can reliably identify targets with high Pb concentration (e.g. C2) from those with low Pb concentration (e.g. D1). This map can also broadly identify the general location of two medieval smelters at locations at targets C2 and D2. These two approximate locations were confirmed more definitively by several sets of separate survey using 159 *ex situ* measurements [43], which were spatially modelled to give contours of equal Pb concentration by a technique known as kriging (Figure A2.5b).

However, although already fit for purpose in terms of meeting the minimum requirement of measurement uncertainty, the use of (say n-fold) composite measurements within each sampling target, would be predicted to reduce sampling uncertainty and hence measurement uncertainty (by √n) and thus further improve the reliability of the resultant geochemical maps of Pb concentration.

Purpose 2: To achieve the optimal (target) level of measurement uncertainty, the results of the Optimised Uncertainty calculation show that a reduction by a factor of 5.8 (= 786/138) is required to achieve fitness for purpose. The measurement uncertainty is dominated by the contribution from sampling (i.e. 99.7 % sampling uncertainty from previous application of Equation A2.1). This means that the most cost-effective way to reduce measurement

uncertainty is by reducing sampling uncertainty, even though the cost of each sampling (€ 29) is over twice the cost of each analysis (€ 12, Table A2.5).

It is possible to reduce sampling uncertainty by a factor of x, by increasing the sample mass by a factor of x^2, according to sampling theory where s^2 is inversely proportional to mass [3, Equation 6, p24]. To reduce the sampling uncertainty by the required factor of 5.8 to achieve fitness for purpose, would therefore require the taking of 34-fold composite measurements at each location (5.8^2), which is clearly impractical. However, even the use of 4-fold composite measurements (say at the corners of a square meter around the sampling location) would similarly be predicted to reduce the sampling uncertainty by a factor of 2, and hence the measurement uncertainty to around 23 %, which would reduce the overall cost and be closer to achieving fitness for purpose.

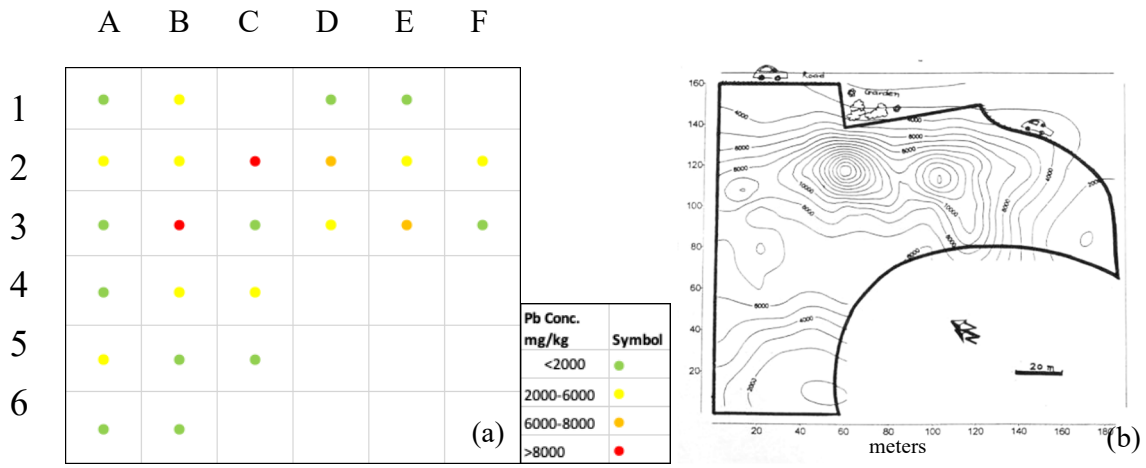

Figure A2.5 - Maps (based upon Figure A2.1) showing the measured Pb concentration from (a) the 24 *in situ* measurements described in this case study, (b) a separate survey using a much large number (159) of *ex situ* measurements [43]. They show that (a) the *in situ* measurements, even with high measurement uncertainty of 55 %, succeeded in locating the parts of the site with high Pb concentrations that are due to the presence of medieval Pb smelters, which were confirmed by (b) the map created from the greater number of *ex situ* measurements

An alternative approach, enabled by the low cost of *in situ* measurement (€ 12), would be to reduce the grid spacing, for example by a factor of 3 from the existing 30 m to 10 m. This higher spatial resolution of surveying could just be applied around the areas of high concentration located in the initial low-density survey. This would increase the measurement costs in those areas, but lower the potential false-remediation costs by a factor a 9 (3^2) and hence lower the consequent costs by a similar factor for the same value of measurement uncertainty [50].

Step 11. Review fitness for purpose of measurement (and analytical) procedure

These particular *in situ* measurements can therefore be shown to be fit for the purpose of geochemical mapping the spatial distribution of the Pb concentration, but *not* for the purpose of classifying contaminated land as being over, or under, a threshold limit of 2000 mg kg^{-1} at this particular site. However, on a different less-contaminated site, say with targets that are all below 500 mg kg^{-1}, *in situ* pXRF measurement values with a similar measurement uncertainty of 55 % may be shown to be fit for the purpose of classification against this threshold concentration of 2000 mg kg^{-1}, which would then be much further above them.

Applicability of the validation

This example shows how it is possible to validate an overall measurement procedure simultaneously, including all of the sampling and analytical steps. In this case the analytical contribution (U'_{anal}) of 3 % has very little effect on the overall measurement uncertainty (but

this is partially due to the very high levels of Pb concentration (mean = ~3000 mg kg^{-1}) which are well above the limit of detection for Pb by this technique, which is 39 mg kg^{-1} [27]. This validation is therefore somewhat site-specific, and this is often the case for application sectors such as contaminated land, where there is often a high degree of variability between different sites. This contrasts with targets and sites that are much more similar, when the validation will be more generally applicable, as is the case in for nitrate in lettuce (Appendix A, Example A1).

Appendix B – Theoretical basis of Optimised Uncertainty method

The Optimised Uncertainty method is one approach to setting the target measurement uncertainty for a measurement procedure that includes sampling as well as chemical analysis. It is briefly discussed in Section 2.3.2, Step 8c, and applied in Examples A1 and A2, both under Step 8. The Optimised Uncertainty method enables the setting of one fitness for purpose criterion that can be used to validate the overall measurement procedure. It can also optimise the relative proportions of measurement uncertainty that are generated by both the sampling and analytical procedures.

The Optimised Uncertainty method compares the effect of varying the measurement uncertainty against estimated total costs. These costs are not just the costs of taking the sample and of making the chemical analysis, but also include the costs that may arise from misclassification of the sampling target. These 'consequence costs' can arise due to the effect of the measurement uncertainty on the compliance/conformity assessment decisions. For the Example A1, if a batch of lettuce gives a false positive measurement value (i.e. false non-compliance) for nitrate concentration, then the batch will be rejected erroneously. This misclassification will cost the producer the overall value of that batch. Conversely, if the measurement value gives a false negative decision (i.e. false compliance), a batch of lettuce with a nitrate value above the EU limit will be sold, eventually to the consumers. If these high nitrate concentration values are detected subsequently, then the producer may well be legally liable for the consequences.

There is an optimal level of measurement uncertainty at which the overall cost (i.e. both measurement costs and consequence costs, formally called Expectation of Loss E(L)) is minimized (Figure B1.1). This optimal level of measurement uncertainty can be used as the target uncertainty at which the measurement procedure can be said to produce measurement values that are fit for that purpose.

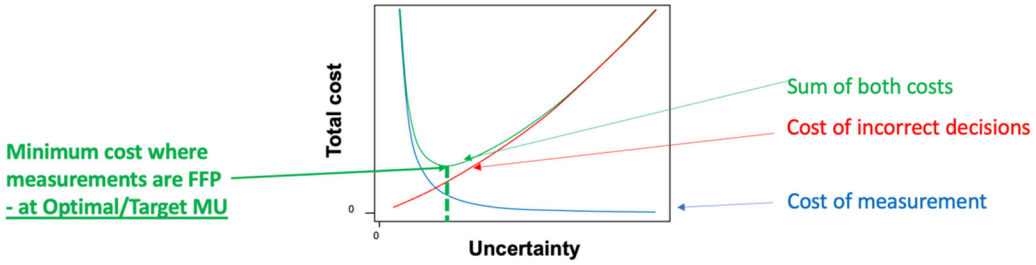

Figure B1.1 - General concept of the Optimised Uncertainty Method. The optimal level of measurement uncertainty is set at the minimal overall cost, including the cost of both the measurement procedure (sampling and analysis) and the potential costs arising from incorrect compliance decisions

Furthermore, a second part of this Optimised Uncertainty approach reveals the relative contributions of the sampling and analysis to both the experimental measurement uncertainty and also to the overall measurement cost. It is possible, therefore, to decide which of these two components must be addressed to change measurement uncertainty to get closer to the target uncertainty and thereby to achieve fitness for purpose.

The equation that gives the total cost (formally expectation of loss) as a function of the measurement uncertainty (green line in Figure B1.1), taken from [19], is:

$$E(L) = C\left[1 - \Phi\left(\frac{\varepsilon}{s_{meas}}\right)\right] + \frac{D}{s_{meas}^2} \qquad \text{Equation B1.1}$$

Where:-

E(L) = Expectation of Loss (the formal term for the total cost)

C = Consequence cost

ε = error limit and $\varepsilon = |T - c_m|$

T = threshold value

c_m = analyte concentration at which the optimisation is made

s_{meas}= standard deviation of the measurement values (i.e. standard uncertainty)

D = cost for unit variance for total measurement process (explanation in Equation B1.4)

Φ = probability that the true concentration will be compliant with the threshold, in the case of a false compliance case, for analyte concentration c_m. By subtracting this probability from 1, the probability that the true concentration is non-compliant is calculated, hence the probability of misclassification. (Φ is given by function NORM.S.DIST in Excel,™)

Calculated input parameters

Using the main input parameters (outline above, e.g. Tables A1.4, A2.5), further parameters can be calculated.

The variable A is the cost for unit variance (s_{smp}^2) for sampling, where L_{smp} is the actual cost of sampling:

$$A = L_{\text{smp}} \times s_{\text{smp}}^2 \qquad \text{Equation B1.2}$$

The variable B is the cost for unit variance (s_{ana}^2) for analysis, where L_{ana} is the actual cost of analysis:

$$B = L_{\text{ana}} \times s_{\text{ana}}^2 \qquad \text{Equation B1.3}$$

The calculations of the optimal levels of these two costs (L'_{smp} and L'_{ana}) are given by Equations B1.7 and B1.8

The cost for unit variance for the total measurement process:

$$D = \left(\sqrt{A} + \sqrt{B}\right)^2 \qquad \text{Equation B1.4}$$

Optimal apportionment of expenditure

Once the optimal value of uncertainty has been determined the optimal values of sampling variance (v'_{smp}) and analytical variance (v'_{ana}) are evaluated from the measurement variance (v'_{meas}). By taking the square root of these variance estimates optimal uncertainties for sampling (s'_{smp}) and analysis (s'_{ana}) can be derived.

The optimum variance of sampling:

$$v'_{\text{smp}} = v'_{\text{meas}} \left[\frac{\sqrt{A}}{(\sqrt{A}+\sqrt{B})}\right] \qquad \text{Equation B1.5}$$

The optimum variance of analysis:

$$v'_{\text{ana}} = v'_{\text{meas}} \left[\frac{\sqrt{B}}{(\sqrt{A}+\sqrt{B})} \right] \qquad \text{Equation B1.6}$$

Subsequently the optimal costs of sampling and chemical analysis, L'_{smp} and L'_{ana} respectively are also computed.

The optimum cost of sampling:

$$L'_{\text{smp}} = \frac{A}{v'_{\text{smp}}} \qquad \text{Equation B1.7}$$

The optimum cost of analysis:

$$L'_{\text{ana}} = \frac{B}{v'_{\text{ana}}} \qquad \text{Equation B1.8}$$

Bibliography

1 H. Cantwell (ed.) Eurachem Guide: The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics, (3rd ed.). Available from http://www.eurachem.org

2 Ellison S L R, Williams A (eds) (2012) Eurachem/CITAC Guide: Quantifying Uncertainty in Analytical Measurement, 3rd edition, ISBN 978-0-948926-30-3. Available from http://www.eurchem.org/

3 Ramsey MH, Ellison SLR., Rostron P (eds.) (2019) Eurachem/EUROLAB/ CITAC/Nordtest/ AMC Guide: Measurement uncertainty arising from sampling: a guide to methods and approach, Second Edition, Eurachem, ISBN 978-0-948926-35-8
http://www.eurachem.org/index.php/publications/guides/musamp

4 ISO 17025:2017 General requirements for the competence of testing and calibration laboratories, International Organization for Standardization, Switzerland. https://www.iso.org

5 Bettencourt da Silva R, Ellison SLR (eds.) Eurachem/CITAC Guide: Assessment of performance and uncertainty in qualitative chemical analysis. First Edition, Eurachem (2021). ISBN 978-0-948926-39-6. Available from https://www.eurachem.org

6 JCGM 200 (2008) International Vocabulary of Metrology – basic and general concepts and associated terms (VIM, 3rd edition). Joint Committee for Guides in Metrology (Sevres). https://www.bipm.org/en/committees/jc/jcgm/publications

7 JCGM 100:2008 Evaluation of measurement data - Guide to the Expression of Uncertainty in Measurement (GUM). Sevres, (2008).

8 Rostron PD, Fearn T, Ramsey MH (2020) Confidence intervals for robust estimates of measurement uncertainty. Accreditation and Quality Assurance: Journal for Quality, Comparability and Reliability in Chemical Measurement. 25, 107-119. https://doi.org/10.1007/s00769-019-01417-4

9 Cordeiro F (Raposo), Ramsey MH (2023) Sampling in the context of conformity assessment. Accreditation and Quality Assurance: Journal for Quality, Comparability and Reliability in Chemical Measurement. 28, 181–182 (2023).
https://doi.org/10.1007/s00769-023-01542-1

10 Commission Regulation (EC) No 152/2009 laying down the methods of sampling and analysis for the official control of feed, Official Journal of the European Union, L54 (2009).

11 Raposo F.C. and Ramsey M.H. (2024) The role of measurement uncertainty in the validation of a measurement procedure. Accreditation and Quality Assurance, Journal for Quality, Comparability and Reliability in Chemical Measurement.
https://doi.org/10.1007/s00769-024-01593-y

12 Ramsey MH, Squire S, Gardner MJ (1999) Synthetic reference sampling target for the estimation of measurement uncertainty. Analyst 124, 11, 1701-1706. https://doi.org/10.1039/A905299B

13 Ramsey MH, Geelhoed B, Damant, AP, Wood, R (2011) Improved evaluation of measurement uncertainty from sampling by inclusion of between-sampler bias using sampling proficiency testing. Analyst, 136 (7), 1313 – 1321. https://doi.org/10.1039/C0AN00705F

14 Ramsey MH, Rostron P (2024) Measurement uncertainty from sampling and its role in validation of measurement procedures. Accreditation and Quality Assurance: Journal for Quality, Comparability and Reliability in Chemical Measurement. https://doi.org/10.1007/s00769-024-01575-0

15 Lyn JA, Ramsey MH, Fussel RJ, Wood R (2003) Measurement uncertainty from physical sample preparation: estimation including systematic error. Analyst, 128 (11), 1391 – 1398.
https://doi.org/10.1039/B307581H

16 Lee J.C. and Ramsey M.H. (2001) Modelling measurement uncertainty as a function of concentration: an example from a contaminated land investigation. Analyst, 126, issue 10, 1784-1791. https://doi.org/10.1039/B104946C

17 Thompson M. (2011) Uncertainty functions, a compact way of summarising or specifying the behaviour of analytical systems. TrAC: Trends in Analytical Chemistry Volume 30, Issue 7, July–August 2011, Pages 1168-1175. *https://doi.org/10.1016/j.trac.2011.03.012*

18 Bettencourt da Silva R, Williams A (Eds), Eurachem/CITAC Guide: Setting and Using Target Uncertainty in Chemical Measurement, (1st ed. 2015). Available from www.eurachem.org

19 Thompson M, Fearn T (1996) What exactly is fitness for purpose in analytical measurement? Analyst, 121, 275–278. https://doi.org/10.1039/AN9962100275

20 Ramsey MH, Lyn JA, Wood R (2001) Optimised uncertainty at minimum overall cost to achieve fitness-for-purpose in food analysis. Analyst, 126, 1777–1783. https://doi.org/10.1039/B104285H

21 Ramsey MH, Taylor PD, Lee JC (2002) Optimized contaminated land investigation at minimum overall cost to achieve fitness-for-purpose. Journal of Environmental Monitoring, 4 (5), 809–814. https://doi.org/10.1039/B203096A

22 Rostron PD, Fearn T, and Ramsey MH (2022) Comparing Uncertainties – Are they really different? Accreditation and Quality Assurance: Journal for Quality, Comparability and Reliability in Chemical Measurement, 27, 133–142. https://doi.org/10.1007/s00769-022-01501-2

23 Lyn JA, Palestra IM, Ramsey MH, Damant AP, 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. https://doi.org/10.1007/s00769-006-0239-0

24 Rostron P, Ramsey MH (2017) Quantifying heterogeneity of small test portion masses of geological reference materials by PXRF: implications for uncertainty of reference values. Geostandards and Geoanalytical Research, 41, 3, 359-473. https://doi.org/10.1111/ggr.12162

25 ISO 17075 = EN 17075+A1 (2023) Water Quality - General requirements and performance test procedures for water monitoring equipment - Continuous measuring devices.

26 ISO 20988:2007. Air quality. Guidelines for estimating measurement uncertainty.

27 Argyraki A, Ramsey MH, Potts PJ (1997) Evaluation of portable XRF for in situ measurements of lead on contaminated land. Analyst 122, 743-749. https://doi.org/10.1039/A700746I

28 Thompson M, Ramsey MH (1995) Quality concepts and practices applied to sampling – an exploratory study. Analyst, 120, 261–270. https://doi.org/10.1039/AN9952000261

29 AMC Technical Briefs. The role of accreditation in ensuring sampling quality. Anal. Methods, 2019, 11, 3358 – 3360. https://doi.org/10.1039/C9AY90095K

30 Nordtest sampler certification scheme handbook. NT ENVIR 008. ISSN: 1459—2800 (Edition 2 – 1, Approved 2015-04). https://www.nordtest.info/wp/2015/04/14/nordtest-sampler-certification-scheme-handbook-version-2-1-nt-envir-008/

31 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.

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

33 BS EN 12014-2:1997 Foodstuffs-Determination of nitrate and/or nitrite content; Part 2; HPLC/IC method for the determination of nitrate content of vegetables and vegetable products. BS EN 12014-2. London.

34 Farrington D, Damant AP, Powell K, Ridsdale J, Walker M, Wood, R (2006) A Comparison of the Extraction Methods used in the UK Nitrate Residues Monitoring Program, Journal of the Association of Public Analysts (Online), 34, 1-11.
http://www.apajournal.org.uk/Volume_34-35/JAPA_Vol_34_pg_1-11/2006-0001-0011.pdf

35 AMC Software.
http://www.rsc.org/Membership/Networking/InterestGroups/Analytical/AMC/Software/

36 Barwick VJ, Ellison SLR (1998) Estimating measurement uncertainty using a cause and effect and reconciliation approach Part 2. Measurement uncertainty estimates compared with collaborative trial expectation. Anal. Commun., 1998, 35, 377–383. https://doi.org/10.1039/A806576D

37 Lyn JA, Ramsey MH, Wood, R (2002) Optimised uncertainty in food analysis: application and comparison between four contrasting 'analyte-commodity' combinations, Analyst, 127, 1252 – 1260. https://doi.org/10.1039/B203669J

38 Lyn JA, Ramsey MH, Wood R (2003) Multi-analyte optimization of uncertainty in infant food analysis. Analyst 128:379–388. https://doi.org/10.1039/B212697B

39 Ramsey MH (2020) Challenges for the estimation of uncertainty of measurements made *in situ.* Accreditation and Quality Assurance: Journal for Quality, Comparability and Reliability in Chemical Measurement. 26(4), 183-192. https://doi.org/10.1007/s00769-020-01446-4

40 Potts, P. J, Webb, P. C., Williams-Thorpe, O., and Kilworth, R. Analysis of silicate rocks using field-portable X-ray fluorescence instrumentation incorporating a mercury(II) iodide detector: a preliminary assessment of analytical performance. Analyst, 1995, 120, 1273.
https://doi.org/10.1039/AN9952001273

41 GPS Accuracy, US Government. https://www.gps.gov/systems/gps/performance/accuracy/

42 Ramsey MH, Solomon-Wisdom GO, Argyraki A (2013) Evaluation of *in situ* heterogeneity of elements in solids: implications for analytical geochemistry. Geostandards and Geoanalytical Research, 37, 4, 379-391. https://doi.org/10.1111/j.1751-908X.2013.00236.x

43 Argyraki A (1997) Estimation of measurement uncertainty in the sampling of contaminated land. PhD Thesis, Imperial College, University of London.
https://spiral.imperial.ac.uk/handle/10044/1/8489

44 Ministry of Agriculture, Fisheries and Food (MAFF). The analysis of agricultural materials; London, 1986; ISBN 0112427626

45 ISO/Guide 33:2015(en)Reference materials — Good practice in using reference materials.

46 Ramsey MH, Ellison SLR (2015) Uncertainty Factor: an alternative way to express measurement uncertainty in chemical measurement. Accreditation and Quality Assurance: Journal for Quality, Comparability and Reliability in Chemical Measurement. 20, 2,153-155. DOI: 10.1007/s00769-015-1115-6

47 Analytical Methods Committee (2002) Fitting a linear functional relationship to data with error on both variables, Technical Brief No.10, Royal Society of Chemistry, London.
https://www.rsc.org/membership-and-community/connect-with-others/join-scientific-networks/subject-communities/analytical-science-community/amc/technical-briefs/

48 Ramsey MH, Ellison SLR (2017) Combined uncertainty factor for sampling and analysis. Accreditation and Quality Assurance, 22(4), 187-189. https://doi.org/10.1007/s00769-017-1271-y

49 Ramsey, MH, 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.
https://doi.org/10.1016/0375-6742(92)90046-B

50 Taylor PD, Ramsey MH, Potts PJ (2004) Balancing measurement uncertainty against financial benefits: a comparison of *in situ* and *ex situ* analysis of contaminated land. Environmental Science and Technology 38, 6824-6831.
https://doi.org/10.1021/es049739p

51 UK Department of the Environment: Interdepartmental Committee on the Redevelopment of Contaminated Land (ICRCL); London, 1987.