Eurachem

CITAC
Cooperation on International
Traceability in Analytical Chemistry

EURACHEM / CITAC Guide

Evaluation of measurement uncertainty from in-house precision and recovery data

MUPR 2026

INTENTIONALLY BLANK

«Ce qui est simple est toujours faux.
Ce qui ne l’est pas est inutilisable»
«What is simple is always false;
what is not is unusable»

Paul Valéry,
Mauvaises pensées et autres,
1941–42.

INTENTIONALLY BLANK

Evaluation of measurement uncertainty from in-house precision and recovery data

First Edition (2026)

Editor

Ricardo Bettencourt da Silva
(Faculdade de Ciências da Universidade de Lisboa)

Composition of the Working Group*

Eurachem members	
S. Ellison *Chair*	*LGC, United Kingdom*
B. Magnusson *Secretary*	*Trollboken AB, Sweden*
R. Bettencourt da Silva	*Univ. Lisboa, Portugal*
R. Becker	*BAM, Germany*
A. Brzyski	*Eurachem Poland*
E. Christie	*Eurachem Ireland*
K. Darbinyan	*Eurachem Armenia*
I. Leito	*Univ Tartu, Estonia*
D. Leontiev	*SPhU, Ukraine*
O. Levbarg	*Ukrmetrteststandart, Ukraine*
M. Melzer	*BAM, Germany*
T. Näykki	*SYKE, Finland*
P. Pablo Morillas	*EUROLAB-España, Spain*
O. Pellegrino	*IPQ/DMET, Portugal*
F. Rolle	*INRIM, Italy*
M. Rösslein	*EMPA St. Gallen, Switzerland*
E. Sahlin	*RISE, Sweden*
A. Svaneborg	*Teknologisk Institut, Denmark*
E. Theodorsson	*LIU, Sweden*
P. Thomas	*SCK CEN, Belgium*
A. Williams	*United Kingdom*
P. Yolcu Omeroglu	*Eurachem TR, Turkey*
V. Zonaras	*Eurachem Greece*
CITAC members	
F. Rebello Lourenço	*Univ. São Paulo, Brazil*
F. Pennecchi	*INRIM, Italy*

*At time of document approval

Acknowledgements

This document has been produced by a joint Eurachem/CITAC Working Group with the composition shown (right). The editor is grateful to all these individuals and organisations and to others who have contributed comments, advice and assistance.

Production of this Guide was in part supported by Fundação para a Ciência e a Tecnologia, Portugal.

Citation

This publication should be cited* as:
"R Bettencourt da Silva (ed.) Eurachem/CITAC Guide: Evaluation of measurement uncertainty from in-house precision and recovery data. First Edition, Eurachem (2026). ISBN 978-972-9348-30-3. Available from https://www.eurachem.org"

* *Subject to journal requirements*

Evaluation of measurement uncertainty from
in-house precision and recovery data

English edition

First Edition (2026)

PDF: MUPR_2026_EN_v01.pdf
Printed 2026-02-11

Content

Foreword

After the publication of the first edition of the "Guide to the Expression of Uncertainty in Measurement" (GUM), in 1993, it became evident that it would be challenging to apply the approach proposed in this document directly to measurements in chemistry. The GUM proposes the definition of a measurement model that explicitly describes the impact of all input quantities and effects on the estimated measured value, and the modelling and combination of their uncertainties to quantify the measurement uncertainty. The guide allows for some simplifications, such as the combined quantification of some uncertainty components, but its application examples do not explore such simplifications.

Soon after the publication of the GUM, it was suggested to quantify the measurement uncertainty from measurement reproducibility or data collected during method validation available in laboratories. These approaches aimed to simplify calculations and, more importantly, to allow for assessing contributions to the measurement uncertainty from the effects that are difficult to explicitly include in a measurement model, for instance, as an additive or multiplicative quantity. In chemistry, measurement performance can be affected in a rather unpredictable way by the properties of the tested item, other than the analyte level. For instance, the determination of some pesticide residues in tomatoes can be affected by the maturation stage of the vegetable in a way difficult to model accurately.

The second edition of the Eurachem guide on "Quantifying Uncertainty in Analytical Measurement" from 2000 already describes how to quantify the combined effect of many uncertainty components to handle the complexity of some measurements. Examples A4 and A6 describe some of these ways of evaluating the uncertainty, but the document does not reveal how these evaluation processes should be conducted in other measurement scenarios.

In 2003, Nordtest published the first edition of the "Handbook for Calculation of Measurement Uncertainty in Environmental Laboratories" (the current 4th edition was published in 2017), which describes how to use measurement reproducibility or in-house performance data for measurement uncertainty evaluation. The same approach is used in the ISO 11352 standard. These documents suggest a simplified way of handling systematic effects by increasing the measurement uncertainty. This avoids the need to correct results for bias or recovery (which may be challenging to determine reliably). This option, which affects measured value accuracy, is conservative about measurement uncertainty (i.e. leads to over- rather than underestimation of uncertainty). These guidelines are also popular in other fields because of their simplicity.

This Eurachem/CITAC guide aims to give analysts additional tools for the evaluation of the measurement uncertainty based on measurement precision and recovery data collected in laboratories. This guide is particularly suited for cases where the previously developed guidance does not allow meeting requirements regarding measurement value accuracy and uncertainty. The document can be applied to the most demanding measurement requirements, and is particularly suited for cases where conformity assessment must be performed on accurately measured values and reliable measurement uncertainty.

The guide is written as a tutorial with detailed explanations of the reasoning and limitations of the options, ensuring the document's educational value while offering multiple ways to optimise measurement uncertainty. Application examples illustrate how to apply both the simplest algorithms and the more complex formulas necessary to handle cases where uncertainty needs to be evaluated over a wide concentration range based on the analysis of a large diversity of reference materials.

1. Scope

This guide provides a simplified evaluation of the uncertainty of measurements in chemistry based on recovery and precision studies performed during the in-house validation of a fully developed measurement method[*], or based on data generated during verification of standard methods. The guide describes the measurement process from the problem description, the definition of measurement requirements, the determination of performance characteristics used for the quantification of the measurement uncertainty and the decision on measurement fitness for the intended use.

This guide is applicable where in-house studies for a specific measurement method have provided:

- data on intermediate precision (or within-laboratory reproducibility) across the intended working range[†] of the measurement method;
- information on trueness in the form of the results of recovery studies, either on independent reference materials or from ‘spiking’ studies.

Detailed calculations are provided for the case where recovery is checked for significant departure from 100 %, and include the case where the measured recovery deviates significantly from 100 %. Brief guidance is also given for cases where recovery is checked for compliance with a permitted range of recovery.

Validation studies are assumed to have followed the principles of the Eurachem Guide on the Fitness for Purpose of Analytical Methods [1] or similar guidance.

The guide can be applied to a wide range of measurements in chemistry based on different principles and aimed at answering a diversity of analytical problems, provided the above conditions are met. The principles may be applicable to other fields of measurement that follow the same principles for validation.

2. Introduction

Many relevant socioeconomic decisions are based on measurement results such as the assessment of an individual’s health from clinical analysis, the evaluation of the compliance of a product with its specification or the safety of a foodstuff taking legal limits as the reference. The protection of interests managed by specification or legal limits is only effective if measurements are adequate for the assessment. Measurements fitness for the intended use depends on the magnitude of the measurement uncertainty. If the measurement uncertainty is smaller than a target (i.e. maximum admissible) uncertainty defined by taking the intended use of the measurement into account, the measurement is adequate [2, 3]. A method is valid if it is applicable to an adequate range of the studied concentrations[‡] and list of matrices (e.g. applicable to the determination of the mass

---

[*] According to the International Vocabulary of Metrology [2], the detailed description of the measurement subject to validation is the “procedure” while the “method” is a generic description of the measurement. However, since many relevant documents and the analytical community still use the term “method” for the detailed description of the measurement, this document does not follow the VIM in this specific detail to improve document readability.

[†] According to the International Vocabulary of Metrology [2], the term “interval” should be used instead of “range” because a range would be the difference between the higher and lower interval values. For instance, the interval 0.1 mg L^{-1} to 0.8 mg L^{-1} has a 0.7 mg L^{-1} range. This guide uses the more frequently used term (i.e. “range”) to specify the concentration values covered by method validation.

[‡] In this document, the term “concentration” mentioned in general discussions is also applicable to the determination of other quantities such as “mass fractions”, “mass concentrations” or “pH”. In examples, the specific studied quantity [2] is mentioned.

fraction of procymidone in foodstuffs of vegetable origin above 1/3 of the maximum admissible value) and produces measurement results with an adequately low uncertainty. The magnitude of the measurement uncertainty and the chosen decision rule determine the risk of an incorrect decision about the compliance or non-compliance of a product [4-6].

Eurachem/CITAC guides have been published [3-5] for evaluating, using, and setting a maximum value for the measurement uncertainty. A guide on the application of the "Guide to the Expression of Uncertainty in Measurement" [7] to measurements in chemistry was developed [5]. Since decisions on whether a sample* complies with limit values depend on the measurement uncertainty, a guide on how to use uncertainty when making compliance decisions was also produced [4]. For cases where no target (maximum) uncertainty is defined in legislation or by the client, another Eurachem/CITAC guide can be used to decide if measurement uncertainty is adequately low [3].

This document aims to give detailed guidance on how to evaluate the measurement uncertainty from performance characteristics determined during the in-house validation of the measurement method that can be used to quantify the combination of random and systematic effects that affect measurements.

This guide complements available guides on the evaluation of the measurement uncertainty based on overall precision and trueness data collected during in-house method validation [8-10] . This document details how to model the variation of the measurement uncertainty with the concentration, how to identify and combine relevant uncertainty components not quantified from precision and trueness studies, and how to manage systematic effects that impact measurements. The approach covered in this guide can be designated "top-down" [11] based on in-house validation data, "supra-analytical"† [12] or "single laboratory validation" [10].

This guide presents the sequence of assessments involved in the evaluation of the measurement uncertainty in a way that allows it to be used as a tutorial.

3. The main stages of decisions based on measurements

The Eurachem and Eurachem/CITAC guides on Fitness for Purpose of Analytical Methods [1], on specifying the Metrological Traceability [13], on Setting the Target Uncertainty [3], on Quantifying the Measurement Uncertainty [5], and on assessing the Compliance of a test item with limit values based on measurement uncertainty [4], describe a sequence of steps aiming at ensuring that measurements are fit for purpose and adequately interpreted. Figure 1 merges the information presented in the various guides on the steps involved when making decisions based on measurements in chemistry. These steps will be discussed in more detail whenever they are not covered in published guidance.

This sequence of steps is triggered by a problem that can be just the characterisation of a sample with limited information about its composition, such as the determination of the composition of a meteorite, the assessment of the compliance of a product with a maximum value or a more complex problem resulting from Research and Development (R&D).

---

* The term "sample" mentioned in general discussions is also applicable to other type of tested items. In some cases, the test item is a specific small object, such as a gold ring, characterised directly without a previous sampling stage.

† The supra-analytical term refers to the use of intralaboratory data more than necessary for the analysis of a specific sample. This designation has the advantage of using a single word.

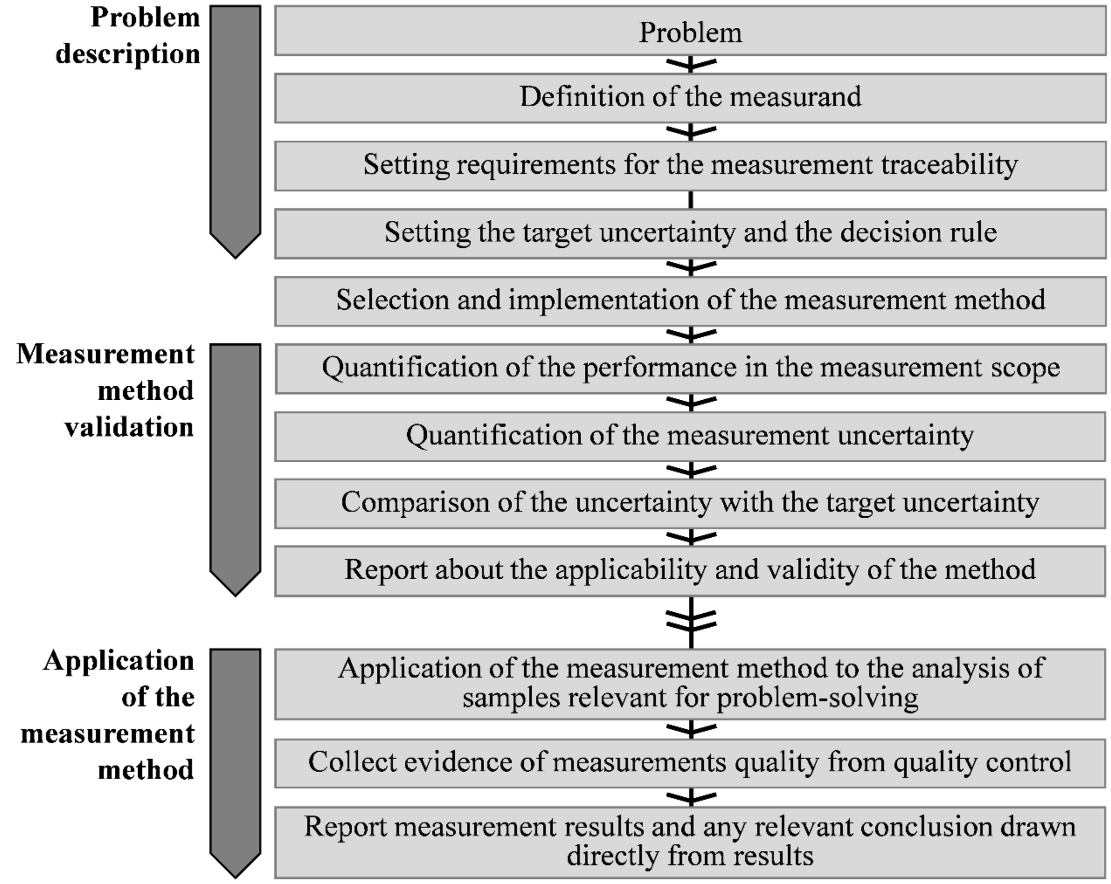

Figure 1: Stages of the measurement process from problem description to problem-solving.

4. Specifying the measurand and the measurement method

The specification of the measurand, i.e. the quantity intended to be measured [2], is not trivial in many cases.

When specifying the measurand, it is necessary to specify the sample (or sampled population), the parameters to be studied, and, in certain cases, the metrological traceability to be achieved.

The matrix of the sample to be analysed should be specified with sufficient detail, depending on method robustness to matrix specificities. For instance, for the measurement of the mass loss of sediment at 103 ºC to 105 ºC, it is not necessary to specify the type of sediment. However, for the measurement of copper mass fraction in sediments by atomic spectroscopy, it is necessary to mention if it is a river or marine sediment due to the inference from high sodium chloride concentration.

When the measurement is performed to solve a specific problem, the analysed sample must be described with all the details necessary to avoid any ambiguity about the information that is being collected. For instance, if the cholesterol in the blood of Mr David is required, it is necessary to specify, in the measurement report, the name or even the identification card number of Mr David and the date and time of blood collection since cholesterol concentration depends on these variables. In laboratories not involved in sample collection, the analysed sample can be adequately described by a code.

The measurand must also clearly specify the parameter to be determined, which is much more than the analyte concentration in many cases. In the determination of chromium (the analyte) level in a sample of an industrial residue, it could be the total chromium, the chromium extractable with *aqua regia* following specific conditions or the chromium extractable with water following a standard method.

The quantity of interest shall be clearly specified, such as a mass fraction (expressed in e.g. mg kg^{-1} unit), an (amount-of-substance) concentration (expressed in e.g. mol L^{-1} unit), a mass concentration (expressed in e.g. mg L^{-1} unit), a pH or an absorbance at a specific wavelength, with all relevant information. In operationally defined measurands, the measurement method must be specified, including the conditions that define the measured quantity (e.g. the temperature, contact time and mixing conditions, proportion of solid sample mass and water and water pH, to estimate the water-soluble fraction of a component of an industrial residue).

Regardless of the type of measurand, the measurement method must be described in adequate detail. The information to be reported includes all the factors that have a relevant impact on the measurement uncertainty, such as balance resolution, particularly for gravimetric measurements, class of volumetric glassware used, particularly for some titrations, type of nebuliser and atomiser in atomic spectroscopy, light path length in molecular or atomic spectrometry, the injection type and detector in gas chromatography, etc. This information is critical to assess which conditions are required to achieve the reported performance.

5. Specifying the basic measurement model

Together with the measurement method, the measurement model [2] used to derive the measured value is provided (i.e. the equation used to calculate the measured value). In certain cases, input quantities should be expressed in terms of further input quantities, e.g. the concentration of a titrant determined from inputs considered from its standardisation or the dried mass of the analytical portion using the mass loss from drying an independent portion of the analysed sample. The analyst should carefully check all the processes specified in the measurement method to be sure that all relevant details are covered. This work is particularly relevant for checking metrological traceability and how the measurand is specified. For instance, after studying in detail how the titrant is standardised, in particular the traceability of the value of the reference used, it becomes clear if measurement results can be traceable to the SI unit mol L^{-1}.

In some cases, the basic measurement model must be updated with terms additional to those considered in measured value calculations to understand how other uncertainty components, not affecting the original input quantities, should be considered (see section 7)

6. Identification of the sources of uncertainty

This stage is particularly demanding in bottom-up uncertainty evaluations [11] but also requires some attention when top-down assessments are performed.

In uncertainty evaluated from overall in-house performance data, uncertainty components are divided into a precision component, trueness-related component also designated bias component by some authors, and additional components. The precision component is quantified by the intermediate precision standard deviation that reflects random effects observed when measurements are performed on different days where, typically, other conditions, such as analysts, equipment and chemicals used, also vary. The trueness-related component quantifies the uncertainty from effects kept constant between days that reflects the impact of some systematic effect on measurements. The additional uncertainty components are all the ones not reflected in precision and trueness-related components but relevant to quantifying the measurement uncertainty. An example of such a component is the sampling uncertainty if relevant for the determination of the measurand and if measurement precision and trueness-related components are not quantified in a way to include random and systematic effects affecting sampling. Section 8.4 describes the quantification of these additional components.

7. Specifying the measurement model

After writing down the basic measurement model, it is often necessary to update it to specify how further effects, not explicitly described in the initial model (e.g. sampling), can affect the measured value and should be combined with the other uncertainty components. These effects can be additive or multiplicative to the measured value or some of its variables.

When effects are additive, such as the possible contamination of analysed samples with the analyte, a variable centred on the mean value of the additive effect (e.g. mean signal of various blank samples*) is subtracted from the original variable (e.g. analysed sample signal)†.

In many cases, the influence quantities affect variables proportionally to the variable value, being managed as a multiplicative unitary factor of the affected variable.

Table 1 presents some common examples where basic measurement models should be updated with additive or multiplicative factors that, in most cases, do not affect the estimated measured value but do contribute to uncertainty evaluation.

Table 1: Examples of multiplicative or additive influence quantities.

Measurand and basic measurement model	Effect implicit in the initial measurement model	Updated measurement model
Mass fraction, w, of total chromium in a soil $w = \frac{V_s \cdot \gamma_s}{m_s \cdot \bar{R}}$	Soil sampling	$w = \frac{V_s \cdot \gamma_s}{m_s \cdot \bar{R}} f$
Mass concentration of ammonium, γ, in a freshwater sample $\gamma = (I_s - a)/b$	Sample contamination	$\gamma = (I_s - \delta - a)/b$

a and b – intercept and slope of analytical instrument calibration;
I_s – sample solution signal;
m_s – mass of the analytical portion ("s" stands for sample);
$\bar{R}$ – Mean analyte recovery;
V_s – volume of the sample solution obtained after the acid digestion of m_s;
γ_s – mass concentration of analyte in the sample solution.
f – unitary multiplicative factor (f = 1) representing the variability from sampling.
δ – additive factor equal to mean blank signal.

---

* Blank samples are samples, originally without the analyte, that can become contaminated.

† If possible, the signals from contaminants should be converted to concentrations to be subtracted from the analysed sample result. Since concentrations cannot be negative, it can be useful to truncate the distribution of blank concentrations below zero.

8. Quantification of the uncertainty components

8.1. Measurement precision

8.1.1. General remarks

For many analytical techniques, measurement precision is a major uncertainty component that can be easily quantified from experimental data.

8.1.2. Determination of measurement repeatability and intermediate precision

For the evaluation of the measurement uncertainty based on in-house measurement performance data, collected from method validation and/or test quality control, two types of precision conditions are relevant: repeatability and intermediate precision conditions [2].

The **repeatability standard deviation**, s_{r}, is determined from results of replicate measurements of the same sample in the same environmental and operational conditions (repeatability condition: tests are performed on the same day by the same analyst and using the same chemicals and equipment). The s_{r} expresses the pure random effects that affect results.

If several (i.e. *N*) $s_{\mathrm{r}(i)}$, determined from the analysis of *N* different sets of test items are available (i.e. $s_{\mathrm{r}(i)}$ determined from n_i replicates, where *i* = 1,…, *N*) and equivalent (see below), these values can be pooled to give a standard deviation, $s_{\mathrm{r(p)}}$, that combines information from the *N* sets of test items using Eq.(1).

$$s_{\mathrm{r(p)}} = \sqrt{\frac{\sum_{i=1}^{N}(n_i - 1)s_{\mathrm{r}(i)}^2}{\sum_{i=1}^{N}(n_i - 1)}} \tag{1}$$

The $s_{\mathrm{r(p)}}$ has $\sum_{i=1}^{N}(n_i - 1)$ degrees of freedom. Since, typically, sample matrix is not expected to affect measurement precision significantly, it is not necessary that these test items cover the scope of the method regarding matrix type. If the matrix type affects measurement precision, precision should be quantified separately for the different types of matrices.

Eq.(1) can be used to pool $s_{\mathrm{r}(i)}$ when the standard deviation is constant, e.g. for concentrations that are close to each other (e.g. where the ratio of any pair of concentration values is between 0.2 and 5). If $s_{\mathrm{r}(i)}$ are determined at high concentrations where the relative standard deviation or coefficient of variance can be considered to be approximately constant, for most purposes, the relative standard deviations, ${s'}_{\mathrm{r}(i)}$, can be pooled (see section Appendix A)[*]. In this case, Eq.(2) is used to estimate a pooled relative repeatability standard deviation, ${s'}_{\mathrm{r(p)}}$.

$${s'}_{\mathrm{r(p)}} = \sqrt{\frac{\sum_{i=1}^{N}(n_i - 1){s'}_{\mathrm{r}(i)}^2}{\sum_{i=1}^{N}(n_i - 1)}} \tag{2}$$

For Eq.(1) and Eq.(2), it is important to guarantee that pooled $s_{\mathrm{r}(i)}$ or ${s'}_{\mathrm{r}(i)}$ are equivalent. This rule assumes that measurement repeatability is constant (or does not vary

---

[*] This approximation can be inadequate for determinations performed well above the limit of quantification where precision requirements may be more stringent.

significantly) between days; i.e. regardless of the day, the agreement between measurements performed on the same day is equivalent. The equivalence of standard deviation can be tested using Cochran's Q or Bartlett's test [14, 15].

Typically, the $s_{\mathrm{r(p)}}$ and/or $s'_{\mathrm{r(p)}}$ are estimated during the validation of the measurement method but can also be obtained from quality control data such as duplicate analysis. Occasionally, target values (i.e. maximum admissible values) are defined for these parameters to decide if the measurement precision is fit for the intended use.*

If the same sample is analysed *n* times, under repeatability conditions, the repeatability standard deviation, $s_{\mathrm{r}}(\bar{x}_{\mathrm{r}}(n))$†, of the mean, $\bar{x}_{\mathrm{r}}(n)$, can be estimated by Eq.(3) [7].

$$s_{\mathrm{r}}(\bar{x}_{\mathrm{r}}(n)) = s_{\mathrm{r}}/\sqrt{n} \tag{3}$$

(i.e., a mean of *n* replicates is $\sqrt{n}$ times more repeatable than single measurements).

The **intermediate precision standard deviation**, s_{I},‡ is obtained as part of the method validation or quality control from results of replicate analysis of samples on different days with relevant operational conditions, such as analyst, available equipment and/or the batch of chemicals used being changed. The s_{I} can be estimated from pooling statistically equivalent $s_{\mathrm{I}(i)}$ or relative intermediate precision standard deviations, $s'_{\mathrm{I}(i)}$, using equations and criteria equivalent to those described to pool measurement repeatability estimates (see Eq.(1) and (2)) [1].

When various equivalent absolute or relative intermediate precision standard deviations, $s_{\mathrm{I}(i)}$ or $s'_{\mathrm{I}(i)}$, are available, Eq.(1) or Eq.(2) can be used to estimate a pooled absolute or relative intermediate precision standard deviation, $s_{\mathrm{I(p)}}$ or $s'_{\mathrm{I(p)}}$, where $s_{\mathrm{r}(i)}$ or $s'_{\mathrm{r}(i)}$ are substituted by $s_{\mathrm{I}(i)}$ or $s'_{\mathrm{I}(i)}$, respectively. The $s_{\mathrm{I}(i)}$ and $s'_{\mathrm{I}(i)}$ are obtained from the in-house method validation and/or quality control data and should be compared with any defined maximum admissible value for these parameters.

The pooling of various $s_{\mathrm{I}(i)}$ or various $s'_{\mathrm{I}(i)}$ to estimate a $s_{\mathrm{I(p)}}$ or a $s'_{\mathrm{I(p)}}$, respectively, with a large number of degrees of freedom (i.e. $\sum_{i=1}^{N}(n_i - 1)$) is particularly relevant when the samples are not stable (e.g. for more than two consecutive days for vitamin D metabolites in human serum) or the test items can only be analysed twice because only a limited amount of the sample is available (e.g. human serum).

When test items are not stable for more than 24h (e.g. vitamin C in orange juice), the intermediate precision can be estimated from the analysis of samples spiked daily at the same level in different days.

Alternatively, s_{r} and s_{I} can be estimated from the Analysis of Variance (ANOVA) of the replicate analysis of the same sample following an experimental design that fits ANOVA requirements [1].

---

* A target value for $s_{\mathrm{r(p)}}$ or $s'_{\mathrm{r(p)}}$ will not impact on the quantified measurement uncertainty. If a conventional maximum range for results from duplicate analysis is used instead, this option can impact the evaluated measurement uncertainty by being more permissible than the statistical control of duplicate results.

† The index of *s* indicates the precision conditions, in this case repeatability, r, and the text between brackets the parameters for which precision is estimated, in this case the mean, $\bar{x}$, of *n* results obtained under repeatability conditions (i.e. $\bar{x}_{\mathrm{r}}(n)$) (See Example 1).

‡ The Nordtest report TR537 [8] uses "within-laboratory reproducibility", s_{Rw}, instead of "intermediate precision", s_{I}.

If s_r and s_I are known and the sample is analysed $n\times p$ times on p different days, where on each day n replicates are collected under repeatability conditions, the mean of the $n\times p$ replicates, $\bar{\bar{x}}(np)$, has an intermediate precision standard deviation, $s_\mathrm{I}\big(\bar{\bar{x}}(np)\big)$, of:

$$s_\mathrm{I}\big(\bar{\bar{x}}(np)\big) = \sqrt{\frac{s_\mathrm{r}^2}{np} + \frac{s_\mathrm{b}^2}{p}} \tag{4}$$

where s_b^2 is the variance attributed to the effect of the day on measurements where $s_\mathrm{b}^2 = s_\mathrm{I}^2 - s_\mathrm{r}^2$. Eq.(4) can be expressed in terms of s_r and s_I values as Eq.(5):

$$s_\mathrm{I}\big(\bar{\bar{x}}(np)\big) = \sqrt{\frac{s_\mathrm{I}^2}{p} + \frac{s_\mathrm{r}^2(1-n)}{np}} \tag{5}$$

For the usual case, where the sample is analysed n times on the same day under repeatability conditions, Eq.(6) is used [16]:

$$s_\mathrm{I}\big(\bar{\bar{x}}(n)\big) = \sqrt{s_\mathrm{b}^2 + \frac{s_\mathrm{r}^2}{n}} = \sqrt{s_\mathrm{I}^2 + \frac{s_\mathrm{r}^2(1-n)}{n}} \tag{6}$$

The repeatability and intermediate precision standard deviations, obtained during method validation, can be subsequently used for routine analysis quality control and measurement uncertainty evaluation.

When using the approach described in this guide for the evaluation of the measurement uncertainty, with some rare exceptions [*], the measurement precision component (incorporating all sources of variability) is the intermediate precision standard deviation estimated by Eq.(5).

Example 1 illustrates how measurement repeatability is used in quality control of routine analysis.

When precision is expected to deteriorate over time, for instance, due to exposure of instrumentation to the samples during their analysis (e.g. dirt accumulation in a mass spectrometer), some analysts prefer to estimate precision uncertainty from a maximum admissible intermediate precision standard deviation. In these cases, instead of using the experimentally determined value, the maximum one is considered. In such cases, the laboratories should occasionally check if the actual precision standard deviation is not greater than this limit value, for instance, from replicate analysis of samples under intermediate precision conditions [17, 18].

Since measurement precision varies with the concentration, Appendix A proposes how information on measurement precision at a few concentration levels can be used to produce conservative precision models for the full analytical range.

---

[*] In some fields, such as pharmaceuticals monitoring, measurement methods are validated and applied to the analysis of the sample on the same day and precision conditions relevant for the evaluation of the measurement uncertainty are repeatability conditions. In this case, precision and mean recovery observed in daily measurements are the relevant values for measurement uncertainty evaluation.

Example 1: Comparison of means obtained under repeatability conditions using a known repeatability standard deviation

On the same day and using the same measurement conditions, duplicate measurements are performed, the results obtained are x_1 and x_2. If $|x_1 - x_2| < 2.8s_\mathrm{r}$, a mean value $\bar{x}_\mathrm{r}(2)$ can be reported and the repeatability standard deviation of this mean $s_\mathrm{r}\left(\bar{x}_\mathrm{r}(2)\right)$ is $s_\mathrm{r}/\sqrt{2}$. Therefore, if two means are produced (i.e. $\bar{x}_\mathrm{r1}(2)$ and $\bar{x}_\mathrm{r2}(2)$) under the same conditions, the absolute value of their difference, $|\bar{x}_\mathrm{r1}(2) - \bar{x}_\mathrm{r2}(2)|$, should be smaller than $2.8s_\mathrm{r}/\sqrt{2}$ at a confidence level of 95 %.

The described quality control of replicate measurements obtained under repeatability conditions (i.e. check of whether $\bar{x}_\mathrm{r1}$ and $\bar{x}_\mathrm{r2}$ are in agreement) requires that s_r is known with a large number of degrees of freedom.

The factor 2.8 is the multiplication of the expansion factor of a standard deviation, s, to 95 % confidence level (approximately 2) with $\sqrt{2}$ ($2.8 = 2\sqrt{2}$) used to convert the standard deviation of a single value, s, into the standard deviation of the difference between two values, s_d ($s_\mathrm{d} = \sqrt{2}s$). [18]

8.1.3. Measurement precision from samples heterogeneity

If samples analysed to estimate the repeatability and intermediate precision are significantly more homogeneous than samples to be subsequently characterised by the method, an additional uncertainty component for sample heterogeneity should be considered.

First, the repeatability standard deviation, $s_\mathrm{r(h)}$, from the analysis of heterogeneous samples should be measured, for instance, from the mean of ranges, $\bar{A}(h)$, of duplicate measurements of heterogeneous samples, where $s_\mathrm{r(h)} = \bar{A}_\mathrm{(h)}/1.128$ [16]. For concentrations where relative precision is approximately constant, relative ranges, $\bar{A'}_\mathrm{(h)}$, should be considered to estimate the relative standard deviation from the analysis of heterogeneous samples, $s'_\mathrm{r(h)}$ ($s'_\mathrm{r(h)} = \bar{A'}_\mathrm{(h)}/1.128$).

If the sample is analysed *np* times on *p* different days where on each day *n* analyses are performed, the mean of *np* values, $\bar{\bar{x}}_\mathrm{(h)}(np)$, is associated with an intermediate precision standard deviation, $s_\mathrm{I}\left(\bar{\bar{x}}_\mathrm{(h)}(np)\right)$, represented by Eq.(7):

$$s_\mathrm{I}\left(\bar{\bar{x}}_\mathrm{(h)}(np)\right) = \sqrt{\frac{s^2_\mathrm{r(h)}}{np} + \frac{s^2_\mathrm{I} - s^2_\mathrm{r}}{p}} \qquad (7)$$

The notation “h” stands for the analysis of heterogeneous samples. Equation 7 is only valid if analysed samples are as heterogeneous as the samples used to determine $s_\mathrm{r(h)}$ which can be checked from duplicate analysis by assessing if the range of duplicate values is smaller than $2.8s_\mathrm{r(h)}$.

The term $(s^2_\mathrm{I} - s^2_\mathrm{r})$, representing s^2_b, is expected to be positive since σ^2_I is not smaller than σ^2_r (the sigma, σ, is used to represent the “true” value of the standard deviation). However, in some cases, repeatability is equivalent to intermediate precision since the between days variance, s^2_b, is not relevant such as in titrations whose end-point detection is the major precision component.

The same concentrations intervals set for constant absolute or relative precision from the analysis of homogeneous samples should be used for the precision from the analysis of heterogeneous samples (see Appendix A). If the repeatability of the analysis of

homogeneous samples is not available, $s_{\mathrm{I}}\left(\bar{\bar{x}}_{(\mathrm{h})}(np)\right)$ can be overestimated by assuming $s_{\mathrm{r}}^2 = 0$ in Eq.(7).
As for the evaluation of the precision uncertainty, the precision of the analysis of heterogeneous samples can be quantified from a target value instead of using the experimentally quantified one. This methodology requires frequent checks if samples are more heterogeneous than a maximum admissible heterogeneity [19].

8.2. Trueness-related uncertainty

8.2.1. General remarks

This section discusses which test items should be analysed for trueness assessment (section 8.2.2), how trueness-related uncertainty should be evaluated considering the analysed test items (section 8.2.3 and, sections 8.2.4 and 8.2.5 for two types of reference test items) and how to evaluate the recovery and decide whether results should be corrected for incomplete recovery (section 8.2.6). Section 8.2.7 discusses the assessment of measurement trueness when recovery varies significantly with the matrix of the test item. Section 8.3 discusses the practice and limitations of increasing the measurement uncertainty to allow for poor recovery, and Section 8.4 describes the evaluation of uncertainty components not covered in precision and recovery components.

8.2.2. Source of a reference value

The measurement precision does not reflect the impact, on the measurement result, of uncertainty effects that are constant under intermediate precision conditions. These effects are, partially or fully, quantified in the trueness-related uncertainty component, designated bias component by some authors [8]. The evaluation of this uncertainty component requires results of measurements of test items that are measured with respect to a reference value, such as certified reference materials, materials from proficiency tests, test items characterised by another method or test items spiked in the laboratory [1]. The reference materials should ideally be analytically equivalent to the analysed test items in terms of analyte speciation, type of matrix and, in some cases, of any relevant binding or interaction between the analyte and the matrix. In some fields, it is possible to produce adequate reference materials by formulation. For example, in the pharmaceutical sector, preparation of test materials with a range of known proportions of active ingredient is possible.
Two types of reference materials can be distinguished when considering the algorithms used to evaluate trueness-related uncertainty:

1) Reference materials independent of the laboratory or spiked blank samples
2) Spiked samples with native analyte

The reference materials independent of the laboratory are reference materials whose reference value is set independently of any measurement performed using the method under evaluation by the laboratory performing method validation, such as certified reference materials, test items from proficiency tests or reference materials characterised by another method. The recovery data of spiked samples with no native concentration are processed and interpreted equivalently to those estimated for reference materials independent of the laboratory. The native concentration is the concentration originally present in the analysed test item, such as nitrate of untreated urban wastewater samples. Surplus proficiency tests materials can provide useful supplementary materials for validation. However, for trueness evaluation, it is important to check the metrological traceability of the assigned value, especially when the value is assigned by consensus of participant results [20, 21].

The assessment of measurement trueness from the analysis of spiked samples with native analyte requires specific algorithms because the sample must be analysed before and after spiking to estimate the recovered added analyte by difference. The spiked value should be large enough to be distinguishable from the native concentration and involve a negligible dilution of the matrix to guarantee that the observed recovery is representative of the typical test items.

8.2.3. General procedure for evaluation of recovery

Assuming that systematic effects are proportional to the studied concentration of the analyte[*], the mean of recovery of the analyte, $\bar{R}$, of the studied item is determined, where a single recovery value, R_i, is the ratio between the observed, c_i, and reference values, C_i ($R_i = c_i/C_i$).

Some authors use the term apparent recovery [22] when the measurement method has no sample pretreatment responsible for deviations between c_i and C_i.

After estimating the mean recovery, in some cases, it is necessary to assess whether 100 % is within the range of the mean recovery and its respective uncertainty to determine whether results need to be corrected for the observed mean recovery. Results can also be corrected for the recovery observed in daily recovery tests performed in parallel with the analysis of the unknown samples, although using few (typically a single) recovery tests for result correction will be more uncertain than using more recovery tests.

The decision of whether a recovery correction should be applied is usually made before the method is validated and implemented in the laboratory, and a procedure for recovery correction (if applied) is included in the documented procedure. The requirement to apply a recovery correction can also be stated in a standardised method or in regulations. The documented procedure would also include whether the correction should be based on a mean recovery obtained at the validation stage, or on a recovery obtained on one or more samples analysed in the same run. Where the documented procedure includes a recovery correction, validation studies can involve the comparison of recovery-corrected values with reference values.

8.2.4. Reference material independent of the laboratory

Reference materials independent of the laboratory are all reference materials that do not rely on previous measurements performed in the laboratory by the assessed method to estimate the recovery. Certified reference materials, test items from proficiency tests and test items characterised by an independent measurement method (e.g. a reference method) are the most common examples of these reference materials.

If a number of these reference materials are used (e.g. the recovery is estimated from tests on seven reference materials, including two certified reference materials and five samples from proficiency tests, $N = 2 + 5$), the mean recovery, $\bar{R}$, is calculated by Eq.(8).[†]

---

[*] This assumption is adequate in most cases, even at low concentration values, since the precision uncertainty component will be dominant and take the variation of precision with the concentration into account. If systematic effects are not proportional but the analytical method is applicable in a rather narrow analytical interval, assessing systematic effects from analyte recovery is also adequate. The nature of systematic effects can be anticipated by the type of sample preparation or measurement method used.

[†] If mean recoveries ($\bar{c}_i/C_i$) combined in the overall recovery, $\bar{R}$, have significantly different uncertainty values, an overall weighted recovery can be considered, which makes the less uncertain ($\bar{c}_i/C_i$) values, more relevant for the estimated overall weighted recovery [41]. For simplicity, equations 8 and 9 are presented assuming this weighing is not relevant.

$$\bar{R} = \frac{\sum_{i=1}^{N} \frac{\bar{c}_i}{C_i}}{N} \qquad (8)$$

where $\bar{c}_i$ is the mean of the results of the analysis of the i^{th} reference material (i = 1 to N) ($\bar{c}_i = \sum c_{ij}/n_i$, where c_{ij} is the estimated result of the j^{th} replicate analysis of the i^{th} reference material and j varies from 1 to n_i) and C_i is the reference value of the i^{th} reference material. The mean recovery is frequently expressed as a percentage.

The standard uncertainty associated with $\bar{R}$ calculated from Eq.(8) is determined from Eq.(9) [23].

$$u_{\bar{R}} = \sqrt{\sum_{i=1}^{N} \left\{ \left(\frac{\bar{c}_i}{C_i}\right)^2 \left[\left(\frac{s(c_i)}{\bar{c}_i\sqrt{n_i}}\right)^2 + \left(\frac{u(C_i)}{C_i}\right)^2 \right] \right\} \Big/ N} \qquad (9)$$

where $s(c_i)$ is the standard deviation of results from replicate measurements of the i^{th} reference material and $u(C_i)$ the standard uncertainty associated with C_i. Eq.(9) results from the application of the law of propagation of uncertainty to Eq.(8) assuming the independence of measurements performed on the different materials. The $u(C_i)$ should be significantly lower than the target uncertainty, typically lower than one-third or one-fifth of this maximum admissible uncertainty. The combination of uncertainty components described in Eq.(9) is recommended if $s(c_i)$ is estimated from at least ten recovery tests to ensure their degrees of freedom is not too low.

If the evaluated recovery uncertainty is to be applied to tests performed on subsequent days, replicates of the analysis of each reference material should be obtained for different days and $s(c_i)$ is the intermediate precision standard deviation. If each of the reference materials is analysed not less than ten times (i.e. $n_i \geq 10$), the $s(c_i)$ can be determined directly from the replicate measurements c_{ij}. However, if the reference material is analysed once or a small number of times (i.e. $n_i < 10$), the $s(c_i)$ should be estimated from models of the variation of measurement precision with the studied concentration as the ones used to evaluate the uncertainty from precision (Appendix A).

Example 2 illustrates the application of Eq.(8) and Eq.(9) to the determination of recovery from the analysis of three reference materials.

Equations (8) to (9) are also applicable to the analysis of spiked blank samples, i.e. samples proven to have no detectable analyte concentrations and spiked above the limit of quantification, where $u(C_i)$ is the standard uncertainty of the spiked concentration C_i.

Example 2: Determination of mean recovery uncertainty for the measurement of ortho-phosphate in fresh water from the analysis of two certified reference materials and one sample from a proficiency test

The following table presents the reference values with respective expanded uncertainty and the mean measured values, intermediate precision standard deviation and number of replicate analysis.

	Measured value		
Reference value, $(C_i \pm 2u(C_i))$ / mg L^{-1} (k = 2; ≈ 95%)	Mean measured value, $\bar{c}_i$/ mg L^{-1}	Standard deviation, $s(c_i)$/ mg L^{-1}	Number of replicate analysis, n_i
Certified reference materials			
(2.24 ± 0.10)	2.31	0.11	12
(1.000 ± 0.050)	1.032	0.090	10
Proficiency test sample			
(1.77 ± 0.23)	1.70	0.14 = 1.77 × 8 %*	1

* From known relative intermediate precision standard deviation above two times the limit of quantification of 8%.

The unweighted mean recovery is:

$$\bar{R} = \frac{\frac{2.31}{2.24} + \frac{1.032}{1.000} + \frac{1.70}{1.77}}{3} = 1.008$$

and the respective standard uncertainty, $u_{\bar{R}}$:

$$u_{\bar{R}} = \sqrt{\sum_{i=1}^{N}\left\{\left(\frac{\bar{c}_i}{C_i}\right)^2\left[\left(\frac{s(c_i)}{\bar{c}_i\sqrt{n_i}}\right)^2+\left(\frac{u(C_i)}{C_i}\right)^2\right]\right\}\Big/ N} =$$

$$= \sqrt{\begin{array}{c}\left\{\left(\frac{2.31}{2.24}\right)^2\left[\left(\frac{0.11}{2.31\sqrt{12}}\right)^2+\left(\frac{0.10/2}{2.24}\right)^2\right]\right\}+ \\ \left\{\left(\frac{1.032}{1.000}\right)^2\left[\left(\frac{0.090}{1.032\sqrt{10}}\right)^2+\left(\frac{0.05/2}{1.000}\right)^2\right]\right\}+ \\ \left\{\left(\frac{1.70}{1.77}\right)^2\left[\left(\frac{0.14}{1.70\sqrt{1}}\right)^2+\left(\frac{0.23/2}{1.77}\right)^2\right]\right\}\end{array}\Bigg/ 3} =$$

$$= \sqrt{\begin{array}{c}0.0007308 + \\ 0.001476 + \\ 0.01015\end{array}\Big/ 3} = 0.03705$$

8.2.5. Spiked sample with native analyte

If recovery is assessed from the analysis of spiked samples with detectable native concentrations, the mean recovery is calculated by Eq.(10).

$$\bar{R} = \frac{\sum_{i=1}^{N} \frac{\bar{c}_i - \bar{c}_{0i}}{c_{+i}}}{N} \tag{10}$$

where $\bar{c}_{0i}$ and $\bar{c}_i$[*] are the means of the measured values of the i^{th} test item (i = 1 to N) before and after spiking, respectively, and c_{+i} the spiked value. The $\bar{c}_{0i}$ is the native concentration.

The standard uncertainty of the mean recovery is given by Eq.(11)[*] [23]:

$$u_{\bar{R}} = \sqrt{\sum_{i=1}^{N} \left\{ \left(\frac{\bar{c}_i - \bar{c}_{0i}}{c_{+i}} \right)^2 \left[\frac{\frac{s^2(c_i)}{n_i} + \frac{s^2(c_{0i})}{m_i}}{(\bar{c}_i - \bar{c}_{0i})^2} + \left(\frac{u(c_{+i})}{c_{+i}} \right)^2 \right] \right\}} \Bigg/ N \tag{11}$$

where $s(c_{0i})$ and $s(c_i)$ are the standard deviations of precision conditions affecting the difference $(\bar{c}_i - \bar{c}_{0i})$ and $u(c_{+i})$ is the standard uncertainty of the spiked concentration c_{+i}.

Eq.(11) is derived from the application of the law of propagation of uncertainty to Eq.(10). In most situations, $\bar{c}_i$ and $\bar{c}_{0i}$ are obtained under repeatability conditions (i.e., the pair $\bar{c}_i$ and $\bar{c}_{0i}$ is quantified in the same run), and the $s(c_i)$ and $s(c_{0i})$ are standard deviations of measurements obtained under repeatability conditions. Whenever $\bar{c}_i$ are $\bar{c}_{0i}$are obtained under repeatability conditions, their difference $(\bar{c}_i - \bar{c}_{0i})$ does not express systematic effects attributed to the daily run making this approach adequate if these are negligible[†]. In a large majority of cases, the n_i and m_i are smaller than ten, and both $s(c_i)$ and $s(c_{0i})$ should be estimated from models of the variation of s_r with c as the ones used to evaluate the uncertainty of the precision (see Appendix A). Ideally, $s(c_i)$ and $s(c_{0i})$ should be known with at least nine degrees of freedom to avoid the need to take the low number of degrees of freedom of these uncertainty components into account on how components are combined and expanded (See Section 10).

If $\bar{c}_i$ and $\bar{c}_{0i}$ are correlated, Eq.(11) can be adapted to the specificities of data correlation by adding covariance terms or following an equivalent methodology.

If spiked samples are measured before and after spiking only once, Eq.(11) is simplified to Eq.(12):

[§§] This formula applies even when the spiked concentration is low compared to the native concentration, indicating a large recovery uncertainty from a poor spiking option.

[*] $\bar{c}_{0i} = \sum c_{0ik}/m_i$; where c_{0ik} is the result of the k^{th} replicate analysis of the i^{th} test item before spiking, k = 1 to m_i, and $\bar{c}_i = \sum c_{ij}/n_i$, where c_{ij} is the result of the j^{th} replicate analysis of the i^{th} test item after spiking, j = 1 to n_i.

[†] In a measurement method involving a sample preparation before an instrumental quantification based on daily calibrations, some systematic effects from the instrumental quantification are not reflected in the estimated mean recovery. If these effects are relevant, the $\bar{c}_i$ and $\bar{c}_{0i}$ can be estimated in different days (i.e. under intermediate precision conditions) or an additional uncertainty component be considered. If regression model used to calibrate the instrumental method of analysis is accurately selected and validated, instrumental quantifications are not affected by relevant systematic effects.

$$u_{\bar{R}} = \sqrt{\sum_{i=1}^{N}\left\{R_i^2\left[\frac{s^2(c_i)+s^2(c_{0i})}{(\bar{c}_i-\bar{c}_{0i})^2}+u'(c_{+i})^2\right]\right\}\Big/N} \tag{12}$$

where R_i is the recovery quantified from the measurement of the i^{th} spiked sample with native concentration ($R_i = (c_i - c_{0i})/c_{+i}$) and $u'(c_{+i})$ the relative standard uncertainty associated with c_{+i}.

Example 3 illustrates the application of Eq.(10) and Eq.(11) to the assessment of recovery from the analysis of two spiked samples with native concentration.

Example 3: Determination of mean recovery uncertainty for the measurement of ortho-phosphate in fresh water from the analysis of two spiked samples with native concentration

Two river water samples, A and B, with native ortho-phosphate concentrations were analysed before and after spiking with different analyte concentrations to determine the mean analyte recovery. Samples were analysed under repeatability (subscript, r) or intermediate precision (subscript, I) conditions (i.e. same day or different days), or once. For the sample analysed once, the considered intermediate precision standard deviation was determined from previously developed models of precision variation with the concentration. The following Table presents the collected data:

Sample	Parameter	Value
A	Mean measured value before spiking, $\bar{c}_{0i}$	35.21 µg/L
	Repeatability standard deviation of measured values before spiking, $s_{\mathrm{r}}(c_{0i})$	1.054 µg/L
	Number of estimated values before spiking, n_{0i} (see Note 1 below)	16
	Mean measured value after spiking, $\bar{c}_i$	66.30 µg/L
	Repeatability standard deviation of measured values after spiking, $s_{\mathrm{r}}(c_i)$	1.405 µg/L
	Number of estimated values after spiking, n_i	16
	Spiked concentration, c_+	30.00 µg/L
	Standard uncertainty of the spiked concentration, $u(c_+)$	0.10 µg/L
	Note 1: Over eight days, the sample pairs before and after spiking was analysed twice and the mean range of daily duplicates used to estimate the repeatability standard deviation. Repeatability conditions are relevant for recovery estimated from daily analysis before and after spiking.	
B	$\bar{c}_{0i}$	55.14 µg/L
	Intermediate precision standard deviation of measured value before spiking, $s_{\mathrm{I}}(c_{0i})$, estimated from previous precision models	1.802 µg/L
	n_{0i}	1
	$\bar{c}_i$	64.02 µg/L
	Intermediate precision of measured values after spiking, $s_{\mathrm{I}}(c_i)$	1.934 µg/L
	Number of estimated values after spiking, n_i	5
	Spiked concentration, c_+	10.000 µg/L
	Standard uncertainty of the spiked concentration, $u(c_+)$	0.040 µg/L
	Note 2: On one day, the sample before spiking was analysed and on the subsequent 5 days the sample spiked at 10 µg/L was analysed. The spiking was too low to distinguish the native and spiked analyte affecting the recovery uncertainty. Usually, samples before and after spiking are analysed on the same day.	

Mean recovery, $\bar{R}$:

$$\bar{R} = \frac{\frac{66.30 - 35.21}{30.00} + \frac{64.02 - 55.14}{10.000}}{2} = \frac{1.036 + 0.888}{2} = 96.2\%$$

If Eq(11) is taken:

$$u_{\bar{R}} = \sqrt{\sum_{i=1}^{N} \left\{ \left(\frac{\bar{c}_i - \bar{c}_{0i}}{c_{+i}} \right)^2 \left[\frac{\frac{s^2(c_i)}{n_i} + \frac{s^2(c_{0i})}{m_i}}{(\bar{c}_i - \bar{c}_{0i})^2} + \left(\frac{u(c_{+i})}{c_{+i}} \right)^2 \right] \right\} \Big/ N} =$$

$$= \sqrt{\begin{array}{l} \left\{ (1.036)^2 \left[\frac{\frac{1.405^2}{16} + \frac{1.054^2}{16}}{(66.30 - 35.21)^2} + \left(\frac{0.10}{30.00} \right)^2 \right] \right\} + \\ \left\{ (0.888)^2 \left[\frac{\frac{1.934^2}{5} + \frac{1.802^2}{1}}{(64.02 - 55.14)^2} + \left(\frac{0.040}{10.000} \right)^2 \right] \right\} \end{array} \Big/ 2} =$$

$$= \sqrt{\begin{array}{c} \{0.0002262\} + \\ \{0.03997\} \end{array} \Big/ 2} = 10.02\%$$

The recovery uncertainty can be reduced to 1.5% if only results from the analysis of sample A are considered.

8.2.6. Evaluation of the mean recovery

8.2.6.1. General

After estimating the mean recovery and the respective standard uncertainty using Eq.(8) to Eq.(12), or after pooling mean recoveries estimated from reference materials, of the same or different types, it is necessary to decide whether the mean recovery value is fit for the intended use. In some analytical fields, measurement is expected to recover all of the analyte, suggesting that the mean recovery should be equivalent to 100 %. However, in other fields, other target or limit values are defined for single or mean recovery observed from single or various recovery tests, respectively. For instance, in pesticide residues analysis in foodstuffs for the official control of food and feed across the European Union, the regulator has defined that the mean recovery from the analysis of spiked samples should be between 70 % and 120 %, while a single recovery test can produce recovery values between 60 % and 140 % [24][*]. In the measurement of aromatic amines in leather according to the ISO 17234-1:2015 Standard [25], the recovery of 4-methoxy-m-phenylenediamine or benzidine spiked in leather should be larger than 20 % or 70 %, respectively[†].

---

[*] In pesticide residues analysis in foodstuffs, results are considered to be fit for the intended use when recovery is within a defined interval. If two measurements are performed with different methods with significantly different performance (e.g. mean recoveries of 70 % and 120 %) and measurements are not corrected for recovery, measurement results will be significantly different. However, in this field, it was decided that laboratories should be able to select the method according to the available resources if the measurement bias is acceptable.

[†] In the analysis of aromatic amines in leather, following the ISO 17234:2015 standard, laboratories use the same method to determine operationally defined measurands which ensures the same parameter is determined. The 20 % to 70 % tolerance aims at checking the correct method implementation, and the recovery is not used in the calculation of the measured value or the measurement uncertainty.

Therefore, the assessment of recovery for reference materials can involve three types of scenarios:

1) Flexible selection of the method if the recovery not significantly different from 100%
2) Flexible selection of the method if recovery is within defined limits
3) Use of a specific method to determine operationally defined measurands

In this section, scenarios (1) and (3) are discussed. Scenario (2) is briefly covered in Appendix C.

8.2.6.2. Flexible selection of the method if the recovery is not significantly different from 100%

In many analytical fields, the measurand relates to the total amount of analyte present, and extraction (if any) is intended to recover all the analyte present. Total lead in soil or nitrite in drinking water are two such examples. In these cases, the laboratory should assess whether the mean recovery is equivalent to 1 (i.e. 100 %) by taking the standard uncertainty of the mean recovery using the condition of Eq.(13):

$$\frac{|1-\bar{R}|}{u_{\bar{R}}} \leq t_{v}^{0.95} \qquad (13)$$

where $\bar{R}$ and $u_{\bar{R}}$ are quantified by using Eq.(8) to Eq.(12) or any valid alternative to these equations, and $t_{v}^{0.95}$ is the *t* value of the Student's *t* distribution for 95 % confidence level and the degrees of freedom, v, associated with $u_{\bar{R}}$. In most cases, v is equivalent to the degrees of freedom associated with the pooled standard deviation used to quantify $u_{\bar{R}}$. When more than ten recovery tests are performed $t_{v}^{0.95}$ equal to 2 can be considered. In Eq.(13), if terms $\bar{R}$ and $u_{\bar{R}}$ are expressed in percentage, the number 1 should be changed for 100.

If the condition of Eq.(13) is met, the quantified mean recovery, $\bar{R}$, may be taken as 1.0 (100 %), and the measured values from the determination on unknown test items do not need to be corrected for recovery. In this case, the relative mean recovery standard uncertainty, $u'_{\bar{R}}$, is calculated using $\bar{R} = 1$ ($u'_{\bar{R}} = u_{\bar{R}}/1$).

However, if the condition of Eq.(13) is not met, the measured concentration, *c*, of an unknown test item (i.e. the initial estimate of the measurand value) should be corrected for recovery by dividing by the mean recovery, $\bar{R}$, to give the corrected measured concentration, c_{C} (Eq.(14)):

$$c_{\mathrm{C}} = c/\bar{R} \qquad (14)$$

In cases where $\bar{R} \neq 1$, $u'_{\bar{R}} = u_{\bar{R}}/\bar{R}$. This correction should only be performed if it is known that the quantified $\bar{R}$ is representative of the performance on the analysis of the unknown test item.

If the mean recovery is estimated from a very large number of recovery tests performed on reference materials with low reference value uncertainty, very small deviations to the ideal 100 % recovery can be distinguished, i.e. considered relevant, since the conditions of Eq.(13) fails. Where the deviation from complete recovery is negligible compared to the combined uncertainty for the measured value, the laboratory may omit the correction of results for significant bias.

As an alternative to correcting results for poor recovery, it may be decided to modify the method to remove the cause of recovery deviations and collect new recovery data, or to report recovery separately, with its uncertainty, so that the end user can perform any correction required. If recovery correction is not allowed, recovery improvement is not possible and separate reporting of recovery is not desirable, an increased recovery uncertainty can be considered (see section 8.3).

In fact, the appropriate action depends on a variety of factors, including any regulations or guidance in force, whether the recovery is adequate for test item, and whether correction leads to a useful reduction in measurement uncertainty.

8.2.6.3. Use of a specific method to determine operationally defined measurands

Most of the methods developed to determine operationally defined measurands are described without mentioning limits for the recovery of the analyte. These methods rely on a strict specification of the measurement conditions, such as extraction temperature and extracting solution composition, that influence the measured concentration (see Example A5 of the Eurachem/CITAC guide [5]). In these cases, the conditions applied define the measured parameter of an operationally defined measurand.

Ideally, conditions that define the determination of operationally defined measurands are set with tolerances, such as temperature ranges, that do not affect the measured concentration significantly if these conditions are applied accurately. These tolerances are frequently established from adequate robustness tests. In most cases, however, other parts of the procedure are not described in detail, and there is a chance that relevant systematic effects can be produced at these stages. For example, while extraction conditions defining the measurand are well specified for *aqua regia* extractable lead, the procedure for determining lead in the extracts is often left to laboratory choice.

Therefore, the measurement uncertainty necessary to assess if operationally defined measurements from two laboratories agree or are compatible [2], or to assess the compliance of the test item with limit values, should take into account two effects that can affect measurement uncertainty:

1) The flexibility of defined measurement conditions
2) Systematic effects affecting measurement steps independent of measurement conditions

For international standard methods for determining operationally defined measurands, the tolerances for relevant measurement conditions are defined from adequate ruggedness tests that prove the variation of conditions within the tolerance does not affect measurement performance significantly. Therefore, in these cases, only the trueness of the steps independent of measurement conditions should be tested and considered in trueness-related uncertainty.[*]

The assessment of systematic effects from the undefined measurement conditions can be performed by the measurement of reference materials for the pertaining stages of the measurement. For instance, for the determination of the copper extracted from marine sediments with *aqua regia* using method A, the apparent recovery of the analyte on atomic spectrometric measurements can be assessed through the analysis of extracts spiked before the spectrometric quantification (i.e. excluding the extraction stage). Eq.(8) to Eq.(12) can be used depending on whether the extract has a relevant native concentration of copper or not.

In methods developed to determine operationally defined measurands, if no test item with native concentration is available, test items can be spiked before the first measurement

---

[*] If the irrelevance of the impact of measurement condition tolerances on measurement results cannot be assumed due the way method was developed, it is advised to evaluate the measurement uncertainty from measurement reproducibility (i.e. between laboratories reproducibility). The use of measurement reproducibility for uncertainty evaluation is described in the ISO 21478 standard [34].

step to allow estimating of the measurement precision[*]. However, the observed recovery of the spiked concentration of analyte must be assessed by considering that deviations from 100 % can be attributed to the defined measurement conditions.

In order to ensure that defined measurement conditions are under control, it is necessary to guarantee that measurements are traceable to an adequate reference (e.g. through the adequate calibration of the instruments used) [13], and that the measurement uncertainty and parameter variability ensure that the measurement conditions are fulfilled. For instance, if a relevant temperature is controlled in an oven with defined tolerance ($T \pm \Delta T$), the assessment of the oven temperature should check whether the maximum and minimum observed temperature in the oven (T_{Max} and T_{Min}, respectively) and the expanded uncertainty of the measured temperatures ($U(T_{Max})$ and $U(T_{Min})$, respectively) guarantee that the actual temperature of the oven is within ($T \pm \Delta T$). The test temperature is under control when ($T_{Max} + U(T_{Max}) \leq T + \Delta T$) and ($T_{Min} - U(T_{Min}) \geq T - \Delta T$).

8.2.7. Variability of the recovery with the analysed test item

In some cases, the recovery of the analyte varies significantly with the type of matrix of the analysed test item due to variability of matrix effects. The impact of the matrix on the recovery observed in a pair of reference materials i and j quantified by single or mean recoveries R_i and R_j, can be assessed by a t-test (Eq.(15)) that checks if the absolute difference of analyte recoveries (left term of Eq.(15)) is not larger than the expanded uncertainty of the difference calculated with a coverage factor k (right term).

$$|R_i - R_j| \leq k\sqrt{u^2(R_i) + u^2(R_j)} \tag{15}$$

where $u(R_i)$ and $u(R_j)$ are the standard uncertainties of recovery values calculated from previously presented Eq.(9) or Eq.(11) to quantify the uncertainty from a single or mean recovery observed from the analysis of a specific test item.

The confidence level of k, P, (typically 95 %) sets the chance $(1 - P)$ of wrongly concluding that R_i and R_j are different when they are actually equivalent. Therefore, Eq.(15) is adequate to compare a pair of recoveries. If three recoveries were estimated (R_1 , R_2 and R_3) their values can be considered equivalent if the multiple paired comparisons indicate equivalence (R_1 vs R_2, R_2 vs R_3 and R_2 vs R_3). However, since three comparisons are performed, the chance of falsely suggesting that recoveries are different becomes larger than $(1 - P)$; this chance should be between $(1 - P)$ and $(1 - P^3)$ depending on the correlation between R_1 , R_2 and R_3.

Therefore, if the three recovery values are different or more recoveries are compared, a χ^2 test should be performed for a statistically sound assessment of recovery differences [26]. This test involves calculating an overall weighted mean recovery, $\bar{\bar{R}}$, from N single or mean recoveries, R_k, (k = 1 to N) using Eq.(16.1) and a χ^2 representing the sum of squares of relative differences between R_k and $\bar{\bar{R}}$ considering the uncertainty of the recovery, $u(R_k)$ (Eq. (16.2)).

$$\bar{\bar{R}} = \sum_{k=1}^{N}\left(\frac{R_k}{u^2(R_k)}\right) \Bigg/ \sum_{k=1}^{N}\left(\frac{1}{u^2(R_k)}\right) \tag{16.1}$$

---

[*] If spiking precision is relevant, the measurement precision will be overestimated.

$$\chi^2 = \sum_{k=1}^{N} \left(\frac{R_k - \bar{\bar{R}}}{u(R_k)} \right)^2 \tag{16.2}$$

If χ^2 is not larger than a critical value, $\chi^2_{0.05,N-1}$, of a chi-squared distribution for $N-1$ degrees of freedom and 95 % confidence level (MS Excel formula: CHISQ.INV(0.95; $N-1$)), it is concluded that observed recoveries are equivalent. If different, the t-test from Eq.(15) can be used to identify the recovery values potentially responsible for the difference that could support revising the scope of the method regarding types of matrices. If the recovery observed for a specific matrix is different from the other matrices, the laboratory may estimate and use the standard uncertainty of the recovery specific to that matrix.

In cases where it is difficult to divide the scope of the method into different groups of matrices for which recovery would be quantified separately, for instance, due to operational reasons, it is suggested to estimate the uncertainty associated with the matrix by quantifying the relative standard deviation of mean recovery values, $s'(\bar{R})$, and to add this term as an additional uncertainty component (See Examples B5 and B6). The addition of $s'(\bar{R})$ is forced by the inability to divide the method scope in matrices associated with different recovery values and involves minor double counting of effects already covered in terms from Eq. (9), (11) and (12) if mean recoveries whose standard deviation is determined are obtained from a larger number of tests and reference values have negligible uncertainty.

If the measurement uncertainty quantified from described models is larger than the target measurement uncertainty and the variability of matrix effects is a relevant uncertainty component, the measurement uncertainty can be reduced by improving the method in order to make it more robust to matrix effects. The separation of the matrix from the analyte by performing a more efficient clean-up stage or the use of the standard addition method are two examples of strategies that reduce the impact of matrix variability on measurement uncertainty.

After performing relevant changes in the measurement method, new measurement performance data needs to be collected to evaluate the measurement uncertainty.

8.3. Recovery uncertainty from mean squared error

Some authors argue that when the cause of low or high recovery is unknown, the original measurement results should not be corrected for recovery, and, instead, the observed measurement error should be used to increase the measurement uncertainty to be reported [8, 9].

Eurolab [10] warn that “Handling of uncorrected bias is a contentious issue, requiring informed judgement. A range of different approaches have been proposed, but a generally accepted procedure has not yet emerged.” This statement remains accurate; though multiple approaches to allowance for uncorrected bias in an uncertainty statement have been proposed, all have individual disadvantages [27, 28]. The approach below follows the general principles of Eurolab [10] and Nordtest [8] guidance; it is relatively simple to implement and is conservative (that is, produces a large uncertainty). General limitations are considered further below.

If the recovery is estimated from the replicate analysis of only one reference material independent of the laboratory or spiked blank sample, the relative measurement error determined by $(1-\bar{R})$ is added to the terms described in Eq.(9) for N equal to 1:

$$u_{\bar{R}} = \left(\frac{\bar{c}}{C}\right)\sqrt{\left(\frac{\bar{c}-C}{C}\right)^2 + \left(\frac{s(c)}{\bar{c}\sqrt{n}}\right)^2 + \left(\frac{u(C)}{C}\right)^2} = \\ = \bar{R}\sqrt{(\bar{R}-1)^2 + \left(\frac{s(c)}{\bar{c}\sqrt{n}}\right)^2 + \left(\frac{u(C)}{C}\right)^2} \quad (17)$$

Whenever different reference materials independent of the laboratory or spiked blank samples are used to assess recovery, it is suggested to use Eq.(18) based on the "mean squared error" (MSE) specified in the equation using the previously presented notation ($\mathrm{MSE} = (\sum(\bar{R}_i - 1)^2/N)^{1/2}$).

$$u_{\bar{R}} = \bar{R}\sqrt{\frac{\sum_{i=1}^{N}(\bar{R}_i - 1)^2}{N} + \frac{\sum_{i=1}^{N}(u(C_i)/C_i)^2}{N}} \quad (18)$$

The MSE term of Eq.(18) depends on $|\bar{R}_i - 1|$ values regardless of whether recoveries are above or below recovery 100 %. Instead of the MSE, the maximum observed $|\bar{R}_i - 1|$ can also be used. The mean of the square of the relative standard uncertainty of the reference values (second term in the square root) is adequate to express this effect if $(u(C_i)/C_i)$ does not vary significantly with the reference materials used. Alternatively, that term can be substituted by the maximum $(u(C_i)/C_i)$ of the reference materials used. Increasing the uncertainty to allow for an uncorrected bias has several general disadvantages that should be considered before such an approach is adopted. These include:

a. Limited consistency with the recommendations of the GUM [7], which strongly recommends applying a correction for recognised, significant, systematic effects;
b. The reported uncertainty interval will generally not be centred on the best (corrected) estimate of the value of the measurand;
c. The resulting standard uncertainty can be much larger than the uncertainty associated with a result that has been corrected for recovery;
d. Risks associated with conformity assessment can be seriously and adversely affected by uncorrected bias when the bias is significant compared to the permitted range. This can require a change to any decision rule used for assessing compliance with a specification.

The option for not correcting for observed relevant recovery deviation accordingly impacts on both the reported value and the uncertainty. The decision to increase measurement uncertainty instead of correcting results for observed mean recovery is only feasible if the reported uncertainty is smaller than a target value.

8.4. Additional uncertainty components

In some cases, the evaluated uncertainty from measurement precision and recovery misses relevant uncertainty components that must be quantified and combined with the former two components.

When measurement precision and trueness are evaluated for the analysis of analytical portions of a laboratory sample and the measurand refers to the characterisation of a sampling target, the sampling uncertainty must be quantified. This component should then be combined with the uncertainty from the precision and trueness-related components of the subsequent measurement steps[*]. Further information on the evaluation of sampling uncertainty is given in the Eurachem guide on the topic [29] and also its use in validation of an entire measurement procedure [30].

Another less obvious example of an additional uncertainty component is the uncertainty of the value of the stock solution used to prepare calibrators of the analytical instrumentation during the validation of the measurement method. Since the collected validation data depends on the value of the stock solution and the quantified measurement uncertainty is to be applied to measurements performed using another lot of the stock solution, the uncertainty of the value of the stock solution should be considered an additional uncertainty component to be combined with s_{I} and $u_{\bar{R}}$. In most cases, this uncertainty component is negligible when compared with s_{I} and $u_{\bar{R}}$; laboratories should, however, confirm that the value of stock or calibration solutions have negligible uncertainties.

9. Combination of the uncertainty components

In top-down uncertainty evaluations, the uncertainty components are combined using the law of propagation of uncertainty since this algorithm is applicable, with no restriction, for this assessment. Other tools can be used to combine uncertainty components, such as the numerical Kragten [5, 31] and Monte Carlo Methods [32]. However, these numerical methods are more useful in bottom-up evaluations of complex measurement models[†].

If the uncertainty model is divided into two intervals, where absolute or relative measurement precision is approximately constant (Interval I and II), typically the combined standard uncertainty, u_{c}, is calculated by Eq.(19) and Eq.(20), respectively.

$$u_{\mathrm{c}}\langle \mathrm{I} \rangle = \sqrt{s_{\mathrm{I}}^{2}\langle \mathrm{I} \rangle + (c u'_{\bar{R}})^{2}} \tag{19}$$

$$u_{\mathrm{c}}\langle \mathrm{II} \rangle = c\sqrt{s'^{2}_{\mathrm{I}}\langle \mathrm{II} \rangle + u'^{2}_{\bar{R}}} \tag{20}$$

where $u_{c}\langle \mathrm{I} \rangle$ and $u_{c}\langle \mathrm{II} \rangle$ are the combined standard uncertainties for Interval I and II, $s_{\mathrm{I}}\langle \mathrm{I} \rangle$ and $s'_{\mathrm{I}}\langle \mathrm{II} \rangle$ are the intermediate precision standard deviation and relative standard deviation for Interval I and II, $u'_{\bar{R}}$ the relative recovery standard uncertainty and c the measured value of the unknown sample. The c is corrected or not for observed recovery. It can be decided to instead determine the relative standard uncertainty ($u'_{\mathrm{c}} = u_{\mathrm{c}}/c$) where in Eq.(19) and Eq.(20) both terms are divided by c.

10. Expansion of the combined standard uncertainty

If precision and trueness-related components are calculated from a reasonable number of experimental tests, typically not less than 10 analyses or 9 degrees of freedom for the intermediate precision and 10 recovery tests, the combined standard uncertainty, u_{c}, can be expanded to approximately 95 % or 99 % confidence level by multiplying by a coverage factor, k, of 2 or 3, producing an expanded uncertainty, U ($U = k u_{\mathrm{c}}$). Observing a student's t table, it can be seen that 3 is associated with a 99.7 % confidence level. However, in pragmatic measurement uncertainty evaluations 3 is considered to expand

---

[*] These steps are frequently designated as analytical steps.

[†] The Monte Carlo Method is useful when the assumption of the linear variation of the measured value with the variation of input variables within their measurement uncertainty is not valid.

standard uncertainties to 99% confidence level due to the difficulty of managing high confidence levels in long Gaussian curve tails.

If one of these components is associated with a low number of degrees of freedom and results need to be reported with expanded uncertainty, the coverage factor can be the Student's t value for the lower number of degrees of freedom of relevant uncertainty components and the considered confidence level.

11. Assessment and optimisation of the measurement uncertainty

The top-down evaluation of the measurement uncertainty is not the best tool to guide the revision of the measurement method aiming at reducing the measurement uncertainty or cost of analysis. Since this evaluation does not require the modelling of each analytical step, it will also not allow the identification of the analytical steps that could be changed to improve performance.

Nevertheless, top-down evaluations allow identification of whether precision, trueness-related or additional components are the major uncertainty component. If the measurement precision is the major component, Eq. (5) allows for planning how replicate analysis can reduce the measurement uncertainty. Regarding recovery, it is possible to learn if the magnitude of this component results from high recovery dispersion, high uncertainty of the reference values, the closeness between native and spiked concentration, etc, from checking the magnitude of terms used to quantify $u'_{\bar{R}}$. This information can be used to find ways to reduce the magnitude of the subcomponent that is most influential.

Equations (21) to (24) describe the formula used to quantify the relative contribution (expressed as a proportion of variance) of precision, χ_{P}, and recovery components, $\chi_{\bar{R}}$, from components combination using Eq.(19) and (20) for Intervals I and II, respectively.

$$\chi_{\mathrm{P}}\langle\mathrm{I}\rangle = \frac{s_{\mathrm{I}}^2\langle\mathrm{I}\rangle}{u_{\mathrm{c}}^2\langle\mathrm{I}\rangle} \tag{21}$$

$$\chi_{\bar{R}}\langle\mathrm{I}\rangle = \frac{(cu'_{\bar{R}})^2}{u_{\mathrm{c}}^2\langle\mathrm{I}\rangle} \tag{22}$$

$$\chi_{\mathrm{P}}\langle\mathrm{II}\rangle = \frac{s_{\mathrm{I}}'^2\langle\mathrm{II}\rangle c^2}{u_{\mathrm{c}}^2\langle\mathrm{II}\rangle} \tag{23}$$

$$\chi_{\bar{R}}\langle\mathrm{II}\rangle = \frac{u'^2_{\bar{R}}\langle\mathrm{II}\rangle c^2}{u_{\mathrm{c}}^2\langle\mathrm{II}\rangle} \tag{24}$$

12. Reporting measurement results

Reporting expanded uncertainty with more than two significant figures is usually not justified. The measured value should be reported with the same decimal places of the expanded uncertainty. The rounding of the measured value and uncertainty should follow the usual rules for technical and scientific assessments when the last figure to be discharged is 5, i.e. keep a previous even value and round up odd numbers. This rule prevents an artificial shift upward of mean values calculated from many results, where numbers before 5 are always rounded up. The fourth and lower significant uncertainty figures are not to be considered in rounding rules. When the second uncertain figure is reported, it can be adequate to use a software rounding strategy different from the one mentioned earlier.

Together with the results expressed as "measured value", c, ± "expanded uncertainty", U, it should be reported their units, \$ (e.g. mg L^{-1}), combined (e.g.: $(x \pm U)\ \$$) or separately (e.g., $x\ \$ \pm U\ \$$), the coverage factor used and the confidence level considered (e.g. $k\ = 2$ for approximately 95% confidence level). A relative expanded uncertainty, U', can be reported, with the measured value being essential to ensure information is clear. The unitless U' should be reported with not more than two significant figures and the measured value, c, reported with decimal places compatible with the reporting of two uncertain figures. In such cases, the decision on how to report the measured value, c, requires determining the absolute expanded uncertainty ($U = U' \cdot c$). In some cases, the standard method being used, regulations, or client instructions specify additional requirements for reporting results, such as listing the method used or particularities of the test or tested item. If the client or standard method requires the reporting of more figures than allowed by the measurement uncertainty, the report should highlight that more significant figures of uncertainty are being reported. Table 2 presents some examples of reporting results with uncertainty.

Table 2: Examples of correct reporting of results with uncertainty.

Measured value, *c* Expanded uncertainty, *U* Expanded relative uncertainty, *U'* Units, *\$*	Result reporting	Observation
c: 0.1559 *U*: 0.0123 *\$*: mg L^{-1}	(0.156 ± 0.012) mg L^{-1}, *k* = 2, conf. level 95 %	Measured value and uncertainty reported in the same units (mg L^{-1})
c: 0.1559 *U'*: 7.9 % *\$*: mg L^{-1}	Either: (0.156 ± 0.012) mg L^{-1}, *k* = 2, conf. level 95 % or 0.156 mg L^{-1} with relative expanded uncertainty of 7.9 %, *k* = 2, conf. level 95 %	Measured value reported in mg L^{-1} and uncertainty as an absolute or relative expanded uncertainty. The measured value should be reported considering the absolute uncertainty.
c: 5364.9 *U*: 235.9 *\$*: µg L^{-1}	Either (5360 ± 240) µg L^{-1}, *k* = 2, conf. level 95 % or (5.36 ± 0.24) mg L^{-1}, *k* = 2, conf. level 95 %	Unit multiple change to facilitate the reporting of just two uncertainty figures. Second uncertainty figure rounded up since the third figure is 5 (the fourth uncertainty figures should not be considered).
c: 5364.9 *U*: 264.2 *\$*: µg L^{-1}	(5.36 ± 0.26) mg L^{-1}, *k* = 2, conf. level 95 %	Kept second uncertainty figure since the third uncertainty figure is 4.

13.References

[1] H. Cantwell (ed.), Eurachem Guide: TheFitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics (3rd ed.), Eurachem, 2025.

[2] JCGM, International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (3rd ed.) (JCGM 200), Sèvres: BIPM, 2012.

[3] R. Bettencourt da Silva and A. Williams (eds.), Eurachem/CITAC Guide: Setting and Using Target Uncertainty in Chemical Measurement, Eurachem, 2015. ISBN 978-989-98723-7-0.

[4] A. Williams and B. Magnusson (eds), Eurachem/CITAC Guide: Use of uncertainty information in compliance assessment (2nd ed.), Eurachem, 2021. ISBN 978-0-948926-38-9.

[5] S. L. R. Ellison and A. Williams (eds.), Eurachem/CITAC Guide: Quantifying Uncertainty in Analytical Measurement (3rd. ed.), Eurachem, 2012. ISBN 978-0-948926-30-3.

[6] JCGM, Evaluation of Measurement Data – The Role of Measurement Uncertainty in Conformity Assessment (JCGM 106), Sèvres: BIPM, 2012.

[7] JCGM, Guide to the Expression of Uncertainty in Measurement (JCGM GUM-1), Sèvres: BIPM, 2008.

[8] B. Magnusson, T. Näykki, H. Hovind, M. Krysell and E. Sahlin, Handbook for calculation of measurement uncertainty in environmental laboratories (NT TR 537 - Edition 4), Oslo: Nordtest, 2017.

[9] ISO, Water quality - Estimation of measurement uncertainty based on validation and quality control data (ISO 11352), Geneva: ISO, 2012.

[10] Eurolab, Measurement uncertainty revisited: Alternative approaches to uncertainty evaluation (Technical Report No. 1/2007), Paris: Eurolab, 2007.

[11] Analytical Methods Committee, "Uncertainty of measurement: implications of its use in analytical science," *Analyst,* vol. 120, pp. 2303-2308, 1995.

[12] R. Bettencourt da Silva, J. R. Santos and M. F. Camões, "A New Terminology for the Approaches to the Quantification of the Measurement Uncertainty," *Accred. Qual. Assur.,* vol. 10, pp. 664-671, 2006.

[13] S. L. R. Ellison and A. Williams (eds.), Eurachem/CITAC Guide: Metrological Traceability in Chemical Measurements - A guide to achieving comparable results in chemical measurement, Eurachem, 2019. ISBN 978-0-948926-34-1.

[14] ISO, Accuracy (trueness and precision) of measurement methods and results - Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method (ISO 5725-2), Geneva: ISO, 2019.

[15] G. W. Snecdecor and W. G. Cochram, Statistical Methods, Hoboken: John Wiley & Sons, Inc., 1989. ISBN 978-0813815619.

[16] C. Palma, V. Morgado and R. Bettencourt da Silva, “Top-down evaluation of matrix effects uncertainty,” *Talanta,* vol. 192, pp. 278-287, 2019.

[17] N. Volovyk, D. Leontiev, V. Petrus, O. Gryzodub and Y. Pidpruzhnykov, “Development of an advanced strategy on the assay method transfer,” *ScienceRise, Pharm. sci.,* vol. 6, pp. 56-67, 2020.

[18] ISO, Accuracy (trueness and precision) of measurement methods and results — Part 6: Use in practice of accuracy values (ISO 5725-6), Geneva: ISO, 1994.

[19] D. Leontiev, V. Petrus, N. Volovyk and O. Gryzodub, “A study of the influence of the test sample inhomogeneity on variability in assay results of desloratadine in film-coated tablets,” *ScienceRise, Pharm. sci.,* vol. 5, pp. 43-51, 2020.

[20] ISO, Statistical methods for use in proficiency testing by interlaboratory comparison (ISO 13528), Geneva: ISO, 2015.

[21] Eurachem, Use of surplus proficiency test items (Eurachem leaflet), Eurachem, 2022.

[22] D. T. Burns, K. Danzer and A. Townshend (eds.), “Use of the terms "Recovery" and "Apparent Recovery" in Analytical Procedures,” *Pure Appl. Chem.,* vol. 74, pp. 2201-2205, 2002.

[23] R. Cordeiro, C. Rosa and R. Bettencourt da Silva, “Measurements recovery evaluation from the analysis of independent reference materials: Analysis of different samples with native quantity spiked at different levels,” *Accred. Qual. Assur.,* vol. 23, pp. 57-71, 2018.

[24] SANTE, Analytical quality control and method validation procedures for pesticide residues analysis in food and feed (SANTE/11312/2021), Brussels: EC, 2021.

[25] ISO, Leather — Chemical tests for the determination of certain azo colorants in dyed leathers — Part 1: Determination of certain aromatic amines derived from azo colorants (ISO 17234-1), Geneva: ISO, 2015.

[26] CCQM, CCQM Guidance note: Estimation of a consensus KCRV and associated Degrees of Equivalence, Sèvres: BIPM, 2013.

[27] B. Magnusson and S. L. R. Ellison, “Treatment of uncorrected measurement bias in uncertainty estimation for chemical measurements.,” *Anal. Bioanal. Chem.,* vol. 390, pp. 201-213, 2008.

[28] G. E. O'Donnell and B. D. Hibbert, “Treatment of bias in estimating measurement uncertainty,” *Analyst,* vol. 130, pp. 721-729, 2005.

[29] M. H. Ramsey, S. L. R. Ellison and R. Rostron (eds.), Eurachem/EUROLAB/CITAC/Nordtest/AMC Guide: Measurement uncertainty arising from sampling, Eurachem, 2019. ISBN 978-0-948926-35-8.

[30] M. H. Ramsey, R. Rostron and F. C. Raposo (eds.), Eurachem/EUROLAB/CITAC/Nordtest/AMC Guide: Validation of Measurement Procedures that Include Sampling, Eurachem, 2024.

[31] J. Kragten, “Calculating standard deviations and confidence intervals with a universally applicable spreadsheet technique,” *Analyst,* vol. 119, pp. 2161-2166, 1994.

[32] JCGM, Evaluation of measurement data — Supplement 1 to the “Guide to the expression of uncertainty in measurement” — Propagation of distributions using a Monte Carlo method (JCGM 101), Sèvres: BIPM, 2008.

[33] QUASIMEME, Quasimeme Laboratory Performance Studies - Programme 2017, Wageningen: Wageningen University, 2017.

[34] ISO, Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty evaluation (ISO 21748), Geneva: ISO, 2017.

[35] A. Williams, “An alternative to the effective number of degrees of freedom,” *Accred. Qual. Assur.,* vol. 4, pp. 14-17, 1999.

[36] OSPAR, JAMP Guidelines for Monitoring Contaminants in Sediments (Agreement Ref. No. 2002-16), London: OSPAR, 2015.

[37] EPA, Acid Digestion of sediments, sludges and soils (Method 3050B), Washington, D.C.: EPA, 1996.

[38] L. A. Currie, “Nomenclature in evaluation of analytical methods, including detection and quantification capabilities (IUPAC Recommendations 1995),” *Pure Appl. Chem.,* vol. 67, pp. 1699-1723, 1995.

[39] I. Kuselman and A. Fajgelj, “IUPAC/CITAC Guide: Selection and use of proficiency testing schemes for a limited number of participants - chemical analytical laboratories (IUPAC Technical Report),” *Pure Appl. Chem.,* vol. 82, pp. 1099-1135, 2010.

[40] T. Pluháček, R. Pechancová, D. Milde and R. Bettencourt da Silva, “Bottom-up uncertainty evaluation of complex measurements from correlated performance data: Determination of total Cr in yeast by ICP-MS after acid digestion,” *Food Chemistry,* vol. 404, p. 134466, 2023.

[41] R. Mateo, C. M. Oliveira, A. M. Díez-Pascual, S. Vera-López, M. P. San Andrés and R. Bettencourt da Silva, “Impact of recovery correction or subjecting calibrators to sample preparation on measurement uncertainty: PAH determinations in waters,” *Talanta,* vol. 207, p. 120274, 2020.

Appendix A. Variation of the measurement precision in the working range

The s_r and s_I estimated at a specific concentration can be inadequate to quantify precision at another concentration. Ideally, precision should be estimated at different concentrations, but laboratories rarely can afford the time or the cost. Therefore, models of the variation of measurement precision with the concentration based on data at few, even one or two, concentration levels should be available. Models based on information from a few concentration levels necessarily result in pessimistic modelling of the performance at concentrations different from the studied ones.

It is often observed in chemical measurements that the relative values s'_r and s'_I decrease with the concentration (black continuous line of Figure A.1), with s_r and s_I approximately constant in a narrow concentration range. For various applications, the narrow concentration range can be defined considering a higher concentration, c_H, not larger than ten times the lower concentration, c_L ($c_\mathrm{H} < 10c_\mathrm{L}$). Whenever the limit of quantification, c_LOQ[*], is quantified under intermediate precision conditions from the standard deviation, $s_\mathrm{I(LOQ)}$, of the analysis of a reference material with a concentration between around $0.5c_\mathrm{LOQ}$ and $5c_\mathrm{LOQ}$[†], the c_LOQ can be used as a reference for worst-case precision models ($c_\mathrm{LOQ} = k_\mathrm{Q} \times s_\mathrm{I(LOQ)}$; where a k_Q equal to 10 is frequently used [A.1]). At the c_LOQ, the $s'_\mathrm{I(LOQ)}(\%) = (s_\mathrm{I(LOQ)}/c_\mathrm{LOQ}) \times 100\ \% = (100/k_\mathrm{Q})\ \%$, above $2c_\mathrm{LOQ}$ or $5c_\mathrm{LOQ}$ the s'_I is not larger than $(100/(k_\mathrm{Q} \times 2))$ or $(100/(k_\mathrm{Q} \times 5))$, respectively, and below $2c_\mathrm{LOQ}$ or $5c_\mathrm{LOQ}$ the $s_\mathrm{I} \cong s_\mathrm{I(LOQ)}$, respectively.

When absolute precision is studied, frequently it is observed that the precision standard deviation increases approximately linearly with the concentrations well above the limit of detection (above around 30[‡] times the limit of detection) but at low concentrations, the precision is approximately constant. In such cases, it can be assumed that absolute precision, s, has a constant absolute component, s_0, and a constant proportional component, s'_1. This gives the model:

$$s = \sqrt{s_0^2 + (c \cdot s'_1)^2} \qquad \text{(A.1)}$$

where c is the measured concentration.

---

[*] The specific LOQ is presented as the italic symbol c_LOQ where *c* stands for concentration at LOQ.

[†] The described approach assumes a prior approximate estimation of c_LOQ from analytical data, such as signal-to-noise ratio or measurement repeatability. If estimated c_LOQ is not between five times more or five times less than the concentration of the reference material used to estimate $s_\mathrm{I(LOQ)}$, tests with another reference material with more adequate concentration level should be performed.

[‡] The 30 multiplier used in an early stage of method performance assessment should be confirmed after gathering relevant data.

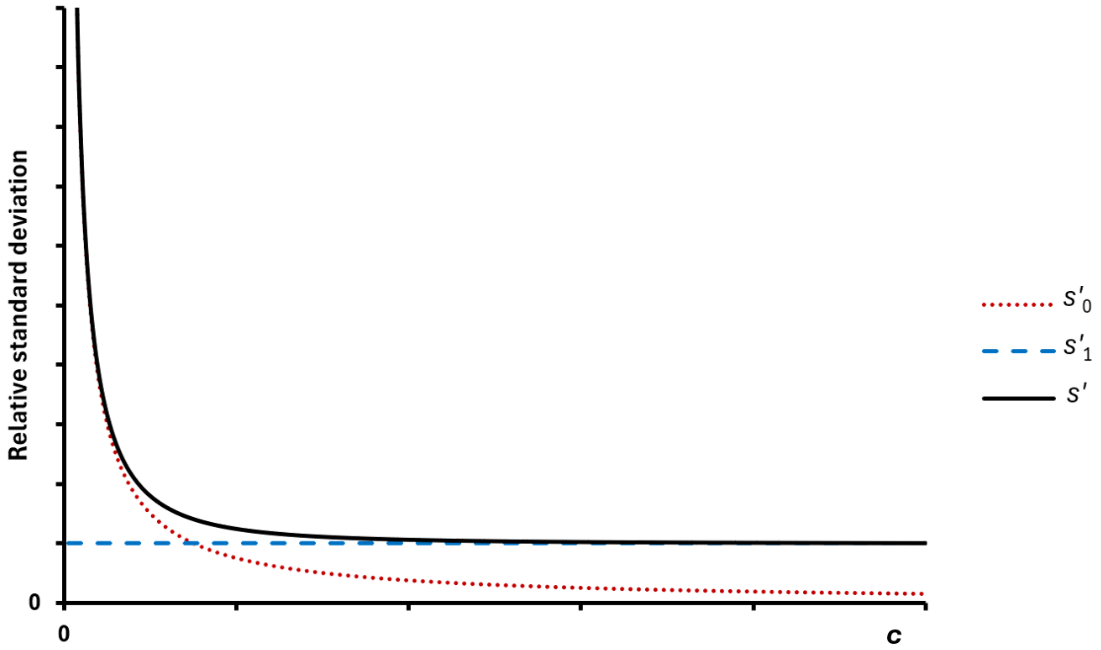

Figure A.1: Typical variation of the precision relative standard deviation, s', with the concentration, c, (—) under repeatability, s'_{r}, or intermediate precision conditions, s'_{I}, ($s' = s'_{\mathrm{I}}$ or $s' = s'_{\mathrm{r}}$) results from the combination of contributions from two components: a constant absolute standard deviation (decreasing relative standard deviation, s'_0, in the figure) (···) and a constant relative standard deviation (---), s'_1.

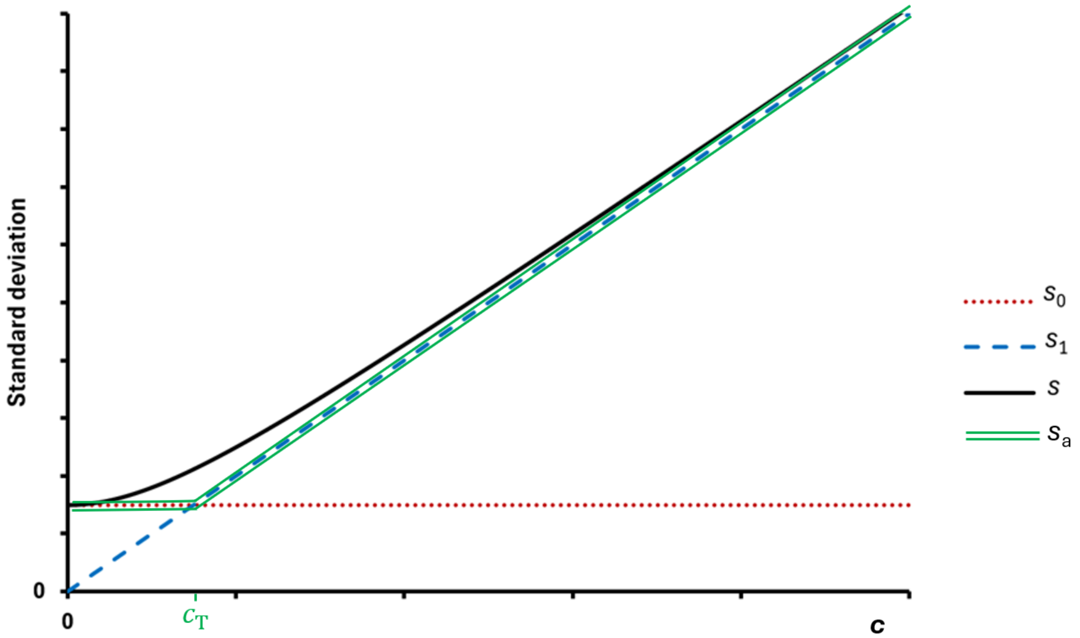

Figure A.2: Typical variation of the intermediate precision, s_{I}, or repeatability, s_{r}, standard deviation, s, ($s = s_{\mathrm{I}}$ or $s = s_{\mathrm{r}}$) (—) with the concentration resulting from the combined contribution of two components: a constant absolute standard deviation, s_0, (···) and a constant relative standard deviation, s'_1 ($s_1 = c \cdot s'_1$), (---). The double line, s_{a}, (═) represents an alternative, less conservative precision model where below and above a transition point ($c_{\mathrm{T}} = s_0/s'_1$) the $s_{\mathrm{a}} = s_0$ or $s_{\mathrm{a}} = c \cdot s'_1$, respectively.

Therefore, the variation of s_I with c can be modelled by estimating the intermediate precision at a higher concentration value, c_H, and a lower value, c_L, where the proportional, $s'_{\mathrm{I}\langle\mathrm{H}\rangle}$, and constant, $s_{\mathrm{I}\langle\mathrm{L}\rangle}$, precision components are determined: $s'_1 = s'_{\mathrm{I}\langle\mathrm{H}\rangle} = s_{\mathrm{I}\langle\mathrm{H}\rangle}/c_\mathrm{H}$ and $s_0 = s_{\mathrm{I}\langle\mathrm{L}\rangle}$. This gives the model:

$$s_\mathrm{I} = \sqrt{s_{\mathrm{I}\langle\mathrm{L}\rangle}^2 + \left(c \cdot s_{\mathrm{I}\langle\mathrm{H}\rangle}/c_\mathrm{H}\right)^2} \qquad \text{(A.2)}$$

If c_L is too high and/or c_H is too low, s_I will be overestimated. For many applications, it would be adequate to use c_L not larger than three times the limit of detection, $3c_\mathrm{LOD}$, and a c_H not lower than $30c_\mathrm{LOD}$ but the diversity of how c_LOD can be determined, and how measurement methods perform suggests using these target values cautiously. An alternative precision model can be used based on a transition concentration, c_T, where $s_{\mathrm{I}\langle\mathrm{L}\rangle} = c_\mathrm{T} \cdot s'_{\mathrm{I}\langle\mathrm{H}\rangle}$ ($c_\mathrm{T} = s_{\mathrm{I}\langle\mathrm{L}\rangle}/s'_{\mathrm{I}\langle\mathrm{H}\rangle}$). Below c_T, $s_\mathrm{I} = s_{\mathrm{I}\langle\mathrm{L}\rangle}$ and above c_T, $s_\mathrm{I} = c \cdot s'_{\mathrm{I}\langle\mathrm{H}\rangle}$ (See Figure A.2).

Therefore, the optimal strategies to model the variation of the intermediate precision over a wide range of concentrations are, by increasing order of the resources required and of the accuracy of the model, the following (Modelling 1 to 4):

1) Determine $s_{\mathrm{I}(c_\mathrm{S})}$ at a suitable level, c_S, such as the regulatory limit or typical levels encountered in practice and use $s_{\mathrm{I}(c_\mathrm{S})}$ to overestimate s_I below c_S and $\left(s_{\mathrm{I}(c_\mathrm{S})}/c_\mathrm{S}\right)$ to overestimate s'_I above c_S. A $c_\mathrm{S} = 2c_\mathrm{LOQ}$ can be suitable for determinations between the c_LOQ and around $10c_\mathrm{LOQ}$ if c_LOQ is estimated under intermediate precision conditions*. This approach requires limited experimental data but overestimates precision away from the studied concentration level.

2) Determine $s_{\mathrm{I(p)}}$ below c_S by pooling several values of s_I determined below that level and estimate $s'_{\mathrm{I(p)}}$ above c_S by pooling several s'_I determined above c_S (see section 8.1.2). Before pooling s_I or s'_I, it is advised to assess the equivalence of pooled parameters using F, Cochran's Q or Bartlett's tests at the 95 % confidence level. This approach has the same disadvantages as the previous one, but combining precision estimated from the analysis of various items benefits the representativeness of quantified precision.

3) Determine the s_I close to the limit of detection, $s_{\mathrm{I}\langle\mathrm{L}\rangle}$, and s'_I at a high concentration, $s'_{\mathrm{I}\langle\mathrm{H}\rangle}$, and use the values to define the transition point c_T ($c_\mathrm{T} = s_{\mathrm{I}\langle\mathrm{L}\rangle}/s'_{\mathrm{I}\langle\mathrm{H}\rangle}$)†. Below c_T, $s_\mathrm{I} = s_{\mathrm{I}\langle\mathrm{L}\rangle}$ and above c_T, $s'_\mathrm{I} = s'_{\mathrm{I}\langle\mathrm{H}\rangle}$ (See s_a line of Figure A.2). Alternatively, it can be decided to produce a more complex and pessimistic continuous precision model for s_I at sample concentration c, $s_\mathrm{I}(c)$.

   $$s_\mathrm{I}(c) = \sqrt{{s_{\mathrm{I}\langle\mathrm{L}\rangle}}^2 + c^2 {s'_{\mathrm{I}\langle\mathrm{H}\rangle}}^2} \qquad \text{(A.3)}$$

   In this case, for each sample concentration, c, the respective $s_\mathrm{I}(c)$ must be calculated (See s line of Figure A.2). The $s_{\mathrm{I}\langle\mathrm{L}\rangle}$ and $s'_{\mathrm{I}\langle\mathrm{H}\rangle}$ can be estimated by

---

* The transition point $2c_\mathrm{LOQ}$ should be confirmed after gathering relevant data.

† As a guide, $s_{\mathrm{I}\langle\mathrm{L}\rangle}$ should be determined at the lowest calibration point or at the limit of detection if that is higher, and $s'_{\mathrm{I}\langle\mathrm{H}\rangle}$ should be determined at the upper end of the working range.

pooling equivalent s_I or s'_I at low or high concentration levels, respectively. This approach has the disadvantage of requiring real items with low and high concentrations that can be difficult to get. Alternatively, items spiked with relevant analyte levels can be considered. The assessment of performance at levels different from typical levels encountered in practice can discourage the use of this approach.

4) Determine s_I and s'_I at various concentrations and fit a line to the variation of s_I or s'_I with the concentration (a linear or a non-linear model can be used). [A.2]

The intermediate precision conditions are the precision conditions relevant for most evaluations of the measurement uncertainty. If it is necessary to evaluate the measurement repeatability by using Eq.(5), models of the variation of s_r with c should be built using the c_S or c_T used to model s_I variation with c (the same transition(s) point(s) should be used for s_I and s_r models).

The same general strategies for accommodating variation of intermediate precision with analyte level can be applied to the estimation of the repeatability standard deviation. Therefore, the optimal strategies to model the variation of s_r with c, for the evaluation of the measurement uncertainty, organised in increasing accuracy and the need for resources of the models, are (Modelling 1 to 4):

1) Estimate $s_{\mathrm{r}(c_\mathrm{S})}$ at c_S and use $s_{\mathrm{r}(c_\mathrm{S})}$ to overestimate s_r below c_S and $\left(s_{\mathrm{r}(c_\mathrm{S})}/c_\mathrm{S}\right)$ to overestimate s'_r above c_S.
2) Estimate $s_{\mathrm{r(p)}}$ below c_S by pooling several s_r estimated below that level and estimate $s'_{\mathrm{r(p)}}$ above c_S by pooling several s'_r estimated above c_S. Before pooling s_r or s'_r, it is advised to assess the equivalence of pooled parameters.
3) Estimate s_r at low, $s_{\mathrm{r}\langle\mathrm{L}\rangle}$, and s'_r at high, $s'_{\mathrm{r}\langle\mathrm{H}\rangle}$, concentration levels to be used as constant s_r or constant s'_r below and above the c_T estimated under intermediate precision conditions (see above). Alternatively, a continuous model of the variation of s_r with sample concentration, c, $s_\mathrm{r}(c)$, can be defined:

$$s_\mathrm{r}(c) = \sqrt{{s_{\mathrm{r}\langle\mathrm{L}\rangle}}^2 + c^2 {s'_{\mathrm{r}\langle\mathrm{H}\rangle}}^2} \quad \text{(A.4)}$$

The $s_{\mathrm{r}\langle\mathrm{L}\rangle}$ and $s'_{\mathrm{r}\langle\mathrm{H}\rangle}$ can be obtained by pooling data from different concentration levels.

4) Estimated s_r and s'_r at various concentrations and fit a line to the variation of s_r or s'_r with concentration.*

A.1. References

[A.1] H. Cantwell (ed.), Eurachem Guide: TheFitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics (3rd ed.), Eurachem, 2025.

[A.2] S. L. R. Ellison and A. Williams (eds.), Eurachem/CITAC Guide: Quantifying Uncertainty in Analytical Measurement (3rd. ed.), 3rd Edition ed., Eurachem, 2012. ISBN 978-0-948926-30-3.

[A.3] See Section 7.1.7 in S. L. R. Ellison, V. J. Barwick, T. J. Duguid Farrant, Practical statistics for the analytical scientist. A bench guide, 2nd ed., RSC Publishing, Cambridge, 2009. ISBN 978-0-85404-131-2.

---

* For direct determinations from instrumental methods of analysis calibrated using the Least Squares Model, the dependence of the prediction interval on concentration can guide the modelling of the variation of precision with the concentration [A.2, A.3].

Appendix B. Application examples

This section describes examples of the evaluation of the measurement uncertainty that aims at illustrating how the described calculation methodologies can be applied. Examples do not necessarily present optimal analytical capabilities for solving the specific problems presented. Therefore, some evaluation options are described due to their pedagogical usefulness instead of analytical relevance.

Examples are presented from a most straightforward analytical problem to a more complex one to make applying proposed algorithms to various scenarios easier. The examples specify basic quality control measures of the tests to illustrate how this control can be consistent with the reported measurement uncertainty. Examples also specify which modelling strategy was applied for the variation of the measurement precision with the concentration (Modelling 1 to 4 from Appendix A). Following examples are considered:

B1: Determination of nitrate in drinking water by ion-chromatography – Uncertainty Evaluation I
Assessment of precision at typical nitrate values of analysed samples using Modelling 1 from Appendix A where the recovery is measured using spiked samples without native analyte when the recovery does not vary with the matrix

B2: Determination of nitrate in drinking water by ion-chromatography – Uncertainty Evaluation II
Assessment of precision above and below typical nitrate values of analysed samples using Modelling 2 from Appendix A where the recovery is measured using two sets of spiked samples without native analyte when the recovery does not vary with the matrix

B3: Determination of nitrate in drinking water by ion-chromatography – Uncertainty Evaluation III
Assessment of precision at four analyte intervals by using data from various samples using Modelling 1 and 4 from Appendix A where recovery is measured using spiked samples without native analyte when the recovery does not vary with the matrix

Conclusions from examples B1, B2 and B3

B4. Determination of total As in sediments – Uncertainty evaluation I
Assessment of precision affected by sample heterogeneity using Modelling 1 from Appendix A and recovery estimated from the analysis of a certified reference material

B5. Determination of total As in sediments – Uncertainty evaluation II
Assessment of precision affected by sample heterogeneity using Modelling 1 from Appendix A and recovery estimated from the analysis of a certified reference material and spiked samples with native analyte where recovery varies with the matrix

B6. Determination of total As in sediments – Uncertainty evaluation III
Assessment of precision affected by sample heterogeneity from the analysis of homogenous and heterogeneous samples using Modelling 1 from Appendix A, and recovery estimated from the single analysis of various proficiency test samples

Conclusions from examples B4, B5 and B6

B1: Determination of nitrate in drinking water by ion-chromatography – Uncertainty Evaluation I

Assessment of precision at typical nitrate values of analysed samples using Modelling 1 from Appendix A where the recovery is measured using spiked samples without native analyte when the recovery does not vary with the matrix

B1.1. Analytical method

Direct determination of nitrate in a drinking water sample received in the laboratory by ion-chromatography with conductivity detection from original or diluted water samples not requiring filtration. The ion chromatograph, IC, was calibrated between 0.2 mg L^{-1} and 1.4 mg L^{-1} [B1.1].

If required, the sample solution can be diluted not more than two times using class A volumetric pipettes between 0.5 mL and 10 mL and class A volumetric flasks between 5 mL and 100 mL. Given possible combinations of these volumetric equipment, the dilution factor will have a maximum, worst-case, relative standard uncertainty of 2.2 % [B1.2]. More details about the method used are available elsewhere [B1.1].

B1.2. Stock solutions and sample conservation

The mass concentration of nitrate in drinking waters, preserved at (5 ± 3) ºC, is stable for 48 h [B1.3]. Below -18 ºC, nitrate content in waters is stable for at least eight days [B1.4]. The eight days of nitrate mass concentration stability is useful for estimating intermediate measurement precision during method validation. Nitrate stock solutions prepared in purified water are stable for at least one month [B1.5].

B1.3. Measurement purpose

The analytical method is used to check drinking water compliance with a maximum nitrate mass concentration of 50 mg L^{-1} defined in Council Directive 98/83/EC [B1.6]. This directive designates this maximum value as a "parametric value" identified in this text with the symbol C^{max} for maximum concentration. The assessment of compliance with C^{max} must be determined by using a measurement method which meets the performance requirements defined in this directive and Commission Directive (EU) 2015/1787 [B1.7]. The Council Directive 98/83/EC also defines a maximum value for the sum of nitrate and nitrite mass concentration in the water [B1.6].

B1.4. Measurement requirements

The Council Directive 98/83/EC [B1.6] defines limit values for "trueness", "precision", and "limit of detection" as proportions of C^{max}. Since these terms are not defined according to the International Vocabulary of Metrology [B1.8], definitions should be checked. If VIM's terminology is used, the measurement requirements are:

(1) Maximum mean absolute error of 5 mg L^{-1} (10 %·50 mg L^{-1})

(2) Maximum absolute repeatability or intermediate precision standard deviation of 2.5 mg L^{-1}

(3) Maximum limit of detection of 5 mg L^{-1}.

The Commission Directive (EU) 2015/1787 [B1.7] sets a maximum or target expanded uncertainty for the determination of nitrate in drinking waters of 7.5 mg L^{-1} at a 95 % confidence level, i.e. 15 % of C^{max}. Since measurement mean error, precision, and ultimately uncertainty is expected to vary with the concentration, these requirements should be applicable around C^{max} (typically between $(100\,\% - 15\,\%)C^{\mathrm{max}}$ and

$(100\ \% + 15\ \%)C^{\max}$; i.e. 42.5 mg L^{-1} and 57.5 mg L^{-1}, since the defined target (maximum) relative expanded uncertainty is 15 %).

Analysis cost and duration should be fit for the analysis of many samples with affordable cost and acceptable duration.

Conformity with the $C^{\max}$ is defined without taking the measurement uncertainty into account in a shared-risk approach strategy associated with a probability of false conformity or non-conformity claims of up to 50 % [B1.9, B1.10].

Note: Directives definitions:

Trueness is the systematic error and is the difference between the mean value of the large number of repeated measurements and the true value (This term is further defined in ISO 5725).

Precision is the random error and is usually expressed as the standard deviation (within and between batch) of the spread of results about the mean. Acceptable precision is twice the relative standard deviation (This term is further defined in ISO 5725).

Limit of detection is either: three times the relative within batch standard deviation of a natural sample containing a low concentration of the parameter, or five times the relative within batch standard deviation of a blank sample.

Uncertainty of measurement is a non-negative parameter characterising the dispersion of the quantity values being attributed to a measurand, based on the information used. The performance criterion for measurement uncertainty ($k = 2$) is the percentage of the parametric value stated in the table or better. Measurement uncertainty shall be estimated at the level of the parametric value, unless otherwise specified.

B1.5. Analytical scope

B1.5.i. Matrices

Drinking water

B1.5.ii. Analytical interval

The analytical method is directly applicable to quantifications between 0.2 mg L^{-1} and 1.4 mg L^{-1} of NO_3, where the maximum analysed nitrate mass concentration can be, in theory, extended to 56 g L^{-1} if the sample is diluted 40000 times. High nitrate values are not expected in drinking waters, but the sample dilution allows determining nitrate concentration far above the calibration interval. Ion chromatograph performance varies in its calibration interval, while dilution uncertainty is independent of analyte concentration.

B1.6. Specific feature of the example

This example describes the simplest scenario of measurement performance assessment: (i) the samples analysed are homogeneous, (ii) the intermediate precision is quantified at the typical nitrate value in samples (i.e. 0.4 mg L^{-1}) (Modelling 1 from Appendix A), and (iii) samples without native analyte can be spiked producing reference materials equivalent to real samples. Spiking is adequate to produce reference materials if it can reproduce the speciation and any relevant binding between the analyte and sample matrix.

B1.6.i. Precision uncertainty

Various drinking waters without detectable native analyte were spiked at 0.4 mg L^{-1} of nitrate, c_1, (the typical nitrate value in samples, around two times the limit of quantification, $2c_{\mathrm{LOQ}}$). The c_{LOQ} was estimated from some preliminary precision tests. The spiked sample was analysed on 20 different days. Daily spiking is allowed if spiking precision is negligible given measurement intermediate precision to avoid overestimating measurement precision*. Daily spiking also avoids variation in analyte content with time.

B1.6.ii. Mean analyte recovery uncertainty
The same data used to estimate the precision uncertainty (see section B1.6.i) was also used to determine the mean apparent analyte recovery. Overall uncertainty evaluation will be more reliable if it is obtained from precision and recovery uncertainties that are based on different sets of data.

B1.6.iii. Mean recovery correction
Since the analysed parameter is rational (i.e., not empirical or operationally defined), any significant deviation between the observed mean recovery and the ideal 100 % value should be corrected. Since the determination of nitrate in waters by IC involves direct sample analysis and no relevant matrix effects are expected, any significant recovery deviations are attributable to the lack of linearity of the calibration curve. Since lack of linearity produces different types of mean recovery deviations, the deviation should be avoided by adequately validating the calibration curve instead of correcting results for recovery. After the adequate validation of the calibration curve, the mean apparent recovery should not be significantly different from100 %.

B1.7. Uncertainty calculations and quality control

B1.7.i. Precision uncertainty
Precision was assessed at the concentration level typically observed in samples (i.e. 0.4 mg L^{-1}) that is far below the parametric or maximum permissible mass concentration in drinking water of 50 mg L^{-1} that can be determined after diluting the water sample. Monitoring at low nitrate concentrations allows for assessing conformity with legislation while informing on the level of this nutrient in water that has seasonal variations.

Table B1.1 presents results from the determination of nitrate in various drinking water samples without native analyte after spiking with (0.400 ± 0.018) mg L^{-1} (coverage factor, k, equal to 2 for approximately 95 % confidence level). Drinking waters from various sources were analysed to ensure that the impact of different sample matrices in the determination was considered, although relevant matrix effects are not expected in this determination. Samples were spiked daily, and their results were obtained from daily calibrations using daily diluted calibrators from a 10 mg L^{-1} nitrate stock solution valid for one month. Independent nitrate stock solutions were used to prepare calibrators and spike samples, both used before the expiration date. However, the uncertainty, in particular, the precision of daily spiking, contributes to observed measurement precision. At the bottom of Table B1.1, the mean, $\bar{c}_1$, (0.389 mg L^{-1}), standard deviation, s_1, (0.033 mg L^{-1}), and coefficient of variance, $s_1/\bar{c}_1 = s'_1$, (8.4 %) of results are reported. Since the reference value of the spiked sample is known, the difference between the estimated mean and reference values must be assessed, as it is discussed below in section B.1.7.ii.
As described in Appendix A (Modelling 1), the measurement precision model is divided into two intervals. Since nitrate measurement involves a direct chromatographic analysis whose performance varies with the nitrate concentration, where sample solution can be

---

[*] For instance, if spiking precision has a relative standard deviation of 20 % by measuring small stock solution volumes and the relative standard deviation of measurement intermediate precision, s'_I, is 5 %, the describe approach would estimate a s'_I of 21 % (21 % = $\sqrt{20\ \%^2 + 5\ \%^2}$) instead of the actual 5 % value.

volumetrically diluted with performance independent of the analyte concentration, Intervals I and II are set for the chromatographic calibration interval.

Interval I [0.2 mg L^{-1}, 0.4 mg L^{-1}][*] where the absolute intermediate precision standard deviation, $s_{\mathrm{I}(1)}\langle\mathrm{I}\rangle$, is overestimated by using a value of 0.033 mg L^{-1}. The symbol $s_{\mathrm{I}(1)}\langle\mathrm{I}\rangle$ stands for the intermediate precision of concentrations c_1 estimated for the first type of samples analysed for method validation and $\langle\mathrm{I}\rangle$ for Interval I.

Interval II [0.4 mg L^{-1}, 1.4 mg L^{-1}] where the relative intermediate precision standard deviation, $s'_{\mathrm{I}(1)}\langle\mathrm{II}\rangle$, is overestimated by using a value of 8.4 % ($s'_{\mathrm{I}(1)}\langle\mathrm{II}\rangle = s_{\mathrm{I}(1)}/\bar{c}_1$).

Table B1.1: Results from the measurement of nitrate in twenty, n_1, drinking water samples without detectable nitrate concentration spiked at (0.400 ± 0.018) mg L^{-1} (coverage factor, k, equal to 2 for approximately 95 % confidence level). The samples were analysed on different days. The mean, $\bar{c}_1$, standard deviation, $s_{\mathrm{I}(1)}$, and relative standard deviation, $s_{\mathrm{I}(1)}/\bar{c}_1$, are reported.

Date	Sample	Nitrate, $c_1/\mathrm{mg\ L^{-1}}$
05/01/2022	50/22	0.416
06/01/2022	53/22	0.332
07/01/2022	107/22	0.385
10/01/2022	117/22	0.395
11/01/2022	124/22	0.381
(…)	(…)	(…)
28/01/2022	270/22	0.395
31/01/2022	279/22	0.467
03/02/2022	285/22	0.378
04/02/2022	286/22	0.424
07/02/2022	289/22	0.441
	$\bar{c}_1$	0.3890 mg L^{-1}
	$s_{\mathrm{I}(1)}$	0.0328 mg L^{-1}
	$s_{\mathrm{I}(1)}/\bar{c}_1$	8.43 %
	n_1	20

Figure B1.1 presents the precision model developed.
Precision from the analysis of a sample that requires dilution is estimated by taking the chromatographic analysis precision and sample dilution precision.

Quality control
The measurement repeatability was also determined at 0.4 mg/L, being the observed repeatability standard deviation, s_{r}, of 0.0136 mg L^{-1} from the analysis of samples not requiring dilution. According to the measured precision, duplicate IC measurements below and above 0.4 mg L^{-1} are accepted if their absolute or relative range is lower than 0.038 mg L^{-1} (2.8·0.0136 mg L^{-1}) or 9.5 % (2.8·0.0136/0.4)) for a 95 % confidence level, respectively. If this control is performed for a 99 % confidence level the multiplying factor for determining maximum ranges should be changed from 2.8 to 3.69. This quality

---

[*] Intervals [ ] include their end points unless otherwise specified.

control is consistent with the measured intermediate precision; i.e. it is not more or less permissible than that allowed by the observed measurement precision.

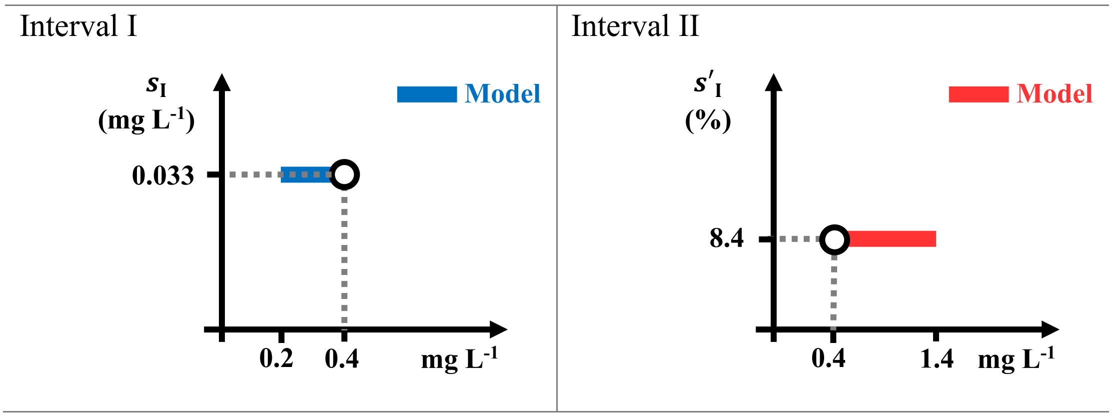

Figure B1.1. Model of intermediate precision variation with the nitrate mass concentration, without sample dilution, measured by IC, developed from results of the analysis of drinking waters without native analyte spiked at 0.4 mg L^{-1} with negligible uncertainty.

B1.7.ii. Recovery uncertainty:

Equations B1.1 and B1.2, based on Equations 8 and 9, were used to determine the mean apparent analyte recovery, $\bar{R}_1$, and the respective standard uncertainty, $u_{\bar{R}1}$.

$$\bar{R}_1 = \frac{\bar{c}_1}{C_1} = \frac{0.389\ \text{mg L}^{-1}}{0.4\ \text{mg L}^{-1}} = 97.2\ \% \tag{B1.1}$$

$$\begin{aligned} u_{\bar{R}1} &= \bar{R}_1 \sqrt{\left(\frac{s_{\text{I}(1)}}{\bar{c}_1\sqrt{n_1}}\right)^2 + \left(\frac{u(C_1)}{C_1}\right)^2} = \\ &= 0.972\sqrt{\left(\frac{0.0328}{0.389\sqrt{20}}\right)^2 + \left(\frac{0.009}{0.4}\right)^2} = 0.0285 \end{aligned} \tag{B1.2}$$

where C_1 is the reference value (i.e. 0.4 mg L^{-1}), n_1 the number of recovery tests and $u(C_1)$ the standard uncertainty of the reference value (i.e., 0.018/2 = 0.009 mg L^{-1}; see Table B.1.1).

The significance of the deviation between the $\bar{R}_1$ and the ideal 1 value is tested by the condition of Eq.(B1.3) based on Eq.(13).

$$\begin{aligned} \frac{|1-\bar{R}_1|}{u_{\bar{R}1}} \leq t_\nu^{0.95} \Leftrightarrow \frac{|1-0.972|}{0.0285} \leq t_{20-1}^{0.95} \Leftrightarrow \\ \Leftrightarrow 0.982 \leq 2.093 \end{aligned} \tag{B1.3}$$

where $t_\nu^{0.95}$ is the *t* value of the Student's t distribution for a 95 % confidence level and ν degrees of freedom ($\nu = n - 1$). Since the condition of Eq.(B1.3) is valid, there is no need to correct results for recovery and the relative standard uncertainty of the mean recovery, $u'_{\bar{R}1}$, is calculated assuming $\bar{R}_1$ is 1 (i.e., $u'_{\bar{R}1} = u_{\bar{R}1}/1 = 0.0285$).

Quality control

The daily calibration curves are checked by analysing a water sample without native analyte spiked at 0.4 mg L^{-1}. If the estimated result produces a daily recovery value between the following limit values, the daily calibration is accepted:

$$L_R^{0.95} = \bar{R}_1 - u_{\bar{R}1} \cdot t_\nu^{0.95} = 0.972 - 0.0848 \cdot 2.093 = 0.794 \qquad \text{(B1.4)}$$

$$H_R^{0.95} = \bar{R}_1 + u_{\bar{R}1} \cdot t_\nu^{0.95} = 0.972 + 0.0848 \cdot 2.093 = 1.15 \qquad \text{(B1.5)}$$

where the $L_R^{0.95}$ and $H_R^{0.95}$ are the lower and upper limits for daily recovery acceptance, and 0.0848 was estimated from Eq.(B1.2) for $n_1 = 1$ since it is defined for single recovery values. These daily recovery limits are set for a 95 % confidence level, being consistent with the estimated recovery uncertainty. The acceptance interval can be defined for a 99 % confidence level by changing $t_\nu^{0.95}$ by the t value of the Student's t distribution for a 99 % confidence level, $t_\nu^{0.99}$ (i.e. $t_{19}^{0.99} = 2.86$) producing new limits: $L_R^{0.99} = 0.729$ and $H_R^{0.99} = 1.21$. The recovery criterion defines a larger tolerance for recovery values below 100 % because the mean recovery is below the ideal recovery.

B1.7.iii. Additional uncertainty components

If the sample solution is diluted before the chromatographic analysis, an additional uncertainty component for the dilution factor must be considered. For a maximum of two sequential dilutions using class A glass pipettes between 0.5 mL and 10 mL and volumetric flasks between 5 mL and 100 mL, gives a maximum relative standard uncertainty, u'_F, associated with the dilution factor F of 0.022 ($u'_F = 2.2$ %) [B1.2]. This component must be considered for diluted samples since precision and recovery were estimated from the analysis of sample solutions not requiring dilution.

Since method validation was conducted over two months using the same nitrate stock solutions to prepare calibrators and spiked samples, and only two dilutions of diluted stock solution of 10 mg L^{-1}, valid for one month, were performed, the uncertainty of both 10 mg L^{-1} solutions were considered as additional sources of uncertainty. The dilution of 1 mL to 100 mL of a nitrate stock solution of (1000 ± 10) mg L^{-1} ($k = 2$), using class A volumetric glassware, produces a Diluted Stock Solution, DSS, with a mass concentration of (10.00 ± 0.19) mg L^{-1} ($k = 2$). Therefore, the relative standard uncertainty, u'_{DSS}, associated with the concentration of the DSS is: $0.0095 = 0.19/2/10$ (i.e. 0.95 %). This component counts twice or once for calibrators and spiked samples prepared from independent 10 mg L^{-1} solutions, if not considered or already considered in $u(C_1)$ from $u'_{\bar{R}1}$ (see Eq. (B1.2)). The Eurachem guide on quantifying the measurement uncertainty [B1.11], namely example A1, discuss how to perform similar calculations. In this specific example, this component is a minor uncertainty component.

B1.7.iv. Uncertainty combination and expansion

The uncertainty components are combined as absolute or relative standard uncertainties depending on whether the original or diluted sample solution analysed by ion chromatography, c_{IA}, is quantified below or above 0.4 mg L^{-1} (Interval I and II, respectively) ("IA" stands for instrumental analysis in this example IC). Since precision, recovery, stock solution value and dilution factor uncertainties were estimated from a large number of tests or are associated with a large number of degrees of freedom, the coverage factors 2 and 3 should be adequate to expand the uncertainty to approximately 95 % and 99 % confidence levels respectively (see section 10).

$$\begin{aligned}U\langle \mathrm{I}\rangle &= 2\sqrt{s^2_{\mathrm{I}(1)}\langle \mathrm{I}\rangle + (c_\mathrm{s}\cdot u'_{\bar{R}1})^2 + 2(c_\mathrm{s}\cdot u'_{\mathrm{DSS}})^2 + (c_\mathrm{s}\cdot u'_F)^2} = \\ &= 2\sqrt{s^2_{\mathrm{I}(1)}\langle \mathrm{I}\rangle + c_\mathrm{S}^2\left(u'^2_{\bar{R}1} + 2u'^2_{\mathrm{DSS}} + u'^2_F\right)} = \\ &= 2\sqrt{0.0328^2 + c_\mathrm{S}^2((0.0285/1)^2 + 2(0.0095)^2 + (0.022)^2)} = \\ &= 2\sqrt{0.0328^2 + c_\mathrm{S}^2(0.00148)}\end{aligned} \tag{B1.6}$$

$$\begin{aligned}U\langle \mathrm{II}\rangle &= 2c_\mathrm{S}\sqrt{s'^2_{\mathrm{I}(1)}\langle \mathrm{II}\rangle + u'^2_{\bar{R}1} + 2u'^2_{\mathrm{DSS}} + u'^2_F} = \\ &= 2c_\mathrm{S}\sqrt{0.0843^2 + (0.0285/1)^2 + 2(0.0095)^2 + (0.022)^2} = \\ &= 2c_\mathrm{S}0.0926 = c_\mathrm{S}0.185\end{aligned} \tag{B1.7}$$

where $U\langle \mathrm{I}\rangle$ and $U\langle \mathrm{II}\rangle$ are expanded uncertainties for a 95 % confidence level applicable below and above 0.4 mg $\mathrm{L^{-1}}$, respectively, and c_S is the measured concentration of the analysed sample ($c_\mathrm{S} = c_\mathrm{IA}\cdot F$). Equations B1.6 and B1.7 are applications of Eq.(19) and (20) with additional uncertainty components considered (i.e. u'_{DSS} and u'_F).

For instance, for a sample with a measured mass concentration of 0.261 mg $\mathrm{L^{-1}}$, not requiring sample dilution (i.e., $F = 1$ and $u'_F = 0$), $U\langle \mathrm{I}\rangle = 2\sqrt{0.0328^2 + 0.261^2(0.000993)} = 0.0676\ \mathrm{mgL^{-1}}$, with the result being reported as (0.261 ± 0.068) mg $\mathrm{L^{-1}}$ for $k = 2$ and approximately 95 % confidence level.

For a sample with a measured mass concentration of 50 mg $\mathrm{L^{-1}}$, the parametric value of the European Legislation [B1.6], requiring 100 times dilution ($F = 100, u'_F = 0.022$, $c_\mathrm{IA} = 0.5\ \mathrm{mg\ L^{-1}}$ and $c_\mathrm{S} = 50\ \mathrm{mg\ L^{-1}}$), $U\langle \mathrm{II}\rangle = 50\cdot 0.185 = 9.26\ \mathrm{mg\ L^{-1}}$, with the result being reported as (50.0 ± 9.3) mg $\mathrm{L^{-1}}$ for $k = 2$ and approximately 95 % confidence level. Interval II is considered since $c_\mathrm{IA} > 0.4\ \mathrm{mg\ L^{-1}}$. This uncertainty is larger than the maximum value defined for determinations close to 50 mg $\mathrm{L^{-1}}$ by European Legislation [B1.6]. Therefore, the described method, including the measurement uncertainty model, is not adequate for the analysis of nitrate concentrations between 42.5 mg $\mathrm{L^{-1}}$ and 57.5 mg $\mathrm{L^{-1}}$. For other mass concentrations of nitrate, the adequacy of the method can be considered acceptable. The measurement uncertainty can be adequately reduced by the replicate analysis of the sample (see section B1.8).

B1.7.v. Uncertainty contributions

Since uncertainty components are combined in a quadratic way as absolute or relative standard uncertainties, uncertainty contributions for measurements at Interval I are calculated as described in Equations B1.8 to B1.11 for the determination of 0.261 mg $\mathrm{L^{-1}}$ of nitrate in a sample not requiring dilution (see section 11).

$$\chi_\mathrm{I}\langle \mathrm{I}\rangle = \frac{s^2_{\mathrm{I}(1)}\langle \mathrm{I}\rangle}{(U\langle \mathrm{I}\rangle/2)^2} = \frac{0.0328^2}{(0.0676/2)^2} = 94.1\ \% \tag{B1.8}$$

$$\chi_{\bar{R}}\langle \mathrm{I}\rangle = \frac{(c_\mathrm{s}\cdot u'_{\bar{R}1})^2}{(U\langle \mathrm{I}\rangle/2)^2} = \frac{(0.261\cdot 0.0285)^2}{(0.0676/2)^2} = 4.8\ \% \tag{B1.9}$$

$$\chi_{\mathrm{DSS}}\langle \mathrm{I}\rangle = \frac{2(c_\mathrm{s}\cdot u'_{\mathrm{DSS}})^2}{(U\langle \mathrm{I}\rangle/2)^2} = \frac{2(0.261\cdot 0.0095)^2}{(0.0676/2)^2} = 1.1\ \% \tag{B1.10}$$

$$\chi_F\langle \mathrm{I}\rangle = \frac{(c_\mathrm{S} \cdot u'_F)^2}{(U\langle \mathrm{I}\rangle/2)^2} = \frac{(0.261 \cdot 0)^2}{(0.0676/2)^2} = 0\ \% \qquad \text{(B1.11)}$$

where $\chi_\mathrm{I}\langle \mathrm{I}\rangle$, $\chi_{\bar{R}}\langle \mathrm{I}\rangle$, $\chi_\mathrm{DSS}\langle \mathrm{I}\rangle$ and $\chi_F\langle \mathrm{I}\rangle$ are the percentage contribution of intermediate precision, mean recovery, both DSS and the dilution factors to the measurement uncertainty. For a c_S equal to 0.261 mg L^{-1}, not requiring sample dilution, $\chi_\mathrm{I}\langle \mathrm{I}\rangle = 94.1\ \%$, $\chi_{\bar{R}}\langle \mathrm{I}\rangle = 4.8\ \%$ and $\chi_\mathrm{DSS}\langle \mathrm{I}\rangle = 1.1\ \%$.

For Interval II, the percentage contribution of the various uncertainty components is calculated from Eq.(B1.12) to (B1.15) for the determination of 50 mg L^{-1} of nitrate in a sample requiring dilution.

$$\chi_\mathrm{I}\langle \mathrm{II}\rangle = \frac{s'^2_{\mathrm{I}(1)}\langle \mathrm{II}\rangle}{(U\langle \mathrm{II}\rangle/(2c_\mathrm{S}))^2} = \frac{0.0843^2}{(9.26/(2 \cdot 50))^2} = 82.9\ \% \qquad \text{(B1.12)}$$

$$\chi_{\bar{R}}\langle \mathrm{II}\rangle = \frac{u'^2_{\bar{R}1}}{(U\langle \mathrm{II}\rangle/(2c_\mathrm{S}))^2} = \frac{0.0285^2}{(9.26/(2 \cdot 50))^2} = 9.5\ \% \qquad \text{(B1.13)}$$

$$\chi_{\bar{R}}\langle \mathrm{II}\rangle = \frac{2u'^2_\mathrm{DSS}}{(U\langle \mathrm{II}\rangle/(2c_\mathrm{S}))^2} = \frac{2 \cdot 0.0095^2}{(9.26/(2 \cdot 50))^2} = 2.1\ \% \qquad \text{(B1.14)}$$

$$\chi_F\langle \mathrm{II}\rangle = \frac{u'^2_F}{(U\langle \mathrm{II}\rangle/(2c_\mathrm{S}))^2} = \frac{0.022^2}{(9.26/(2 \cdot 50))^2} = 5.6\ \% \qquad \text{(B1.15)}$$

For a measured mass fraction of 50.0 mg L^{-1}, $\chi_\mathrm{I}\langle \mathrm{II}\rangle = 82.9\ \%$, $\chi_{\bar{R}}\langle \mathrm{II}\rangle = 9.5\ \%$, $\chi_\mathrm{DSS}\langle \mathrm{II}\rangle = 2.1\ \%$ and $\chi_F\langle \mathrm{II}\rangle = 5.6\ \%$.

B1.8. Uncertainty reduction

Since the intermediate precision is the major component, this component must be reduced to reduce the measurement uncertainty. If Eq.(5) is considered, it can be determined how many replicate analyses of the sample are required to make the measurement uncertainty lower than 7.5 mg L^{-1}, for a measured value of 50 mg L^{-1} or even 57.5 mg L^{-1}. If three replicate measurements of the analysed sample are performed on three different days, the $U\langle \mathrm{I}\rangle$ reduces to 7.1 mg L^{-1} for measured values of 57.5 mg L^{-1}. Eq.(B1.16) describes how this value is obtained.

$$\begin{aligned} U\langle \mathrm{II}\rangle &= 2 \cdot 57.5\sqrt{s'^2_\mathrm{I}\langle \mathrm{II}\rangle/3 + (0.0285/1)^2 + 2(0.0095)^2 + (0.022)^2} = \\ &= 2 \cdot 57.5\sqrt{0.0843^2/3 + 0.001477} = 7.13\ \mathrm{mg\ L^{-1}} \end{aligned} \qquad \text{(B1.16)}$$

Requiring the replicate analysis of the same sample on three different days is inadequate, particularly knowing that the nitrate mass concentration is stable for some refrigerated samples for only 48h.

B1.9. Uncertainty model assumptions and limitations

This measurement uncertainty model assumes the regression model used to build the calibration curve of the IC is adequate, making systematic effects observed in analysis at 0.4 mg L^{-1} equivalent to affecting the whole calibration interval. This model also assumes sample spiking produces samples with known nitrate concentrations equivalent to "real" samples.

The applicability of the uncertainty model can be tested through the described quality control. The models can be further tested through the analysis of other reference materials by checking the metrological compatibility between estimated and reference values. The estimated and reference values with known uncertainty and a large number of degrees of freedom are metrologically compatible for approximately 99 % confidence level, if the following condition is valid.

$$|c_{\mathrm{RM}} - C_{\mathrm{RM}}| \leq 3\sqrt{u(c_{\mathrm{RM}})^2 + u(C_{\mathrm{RM}})^2} = 3u_{\mathrm{d}} \qquad \text{(B1.17)}$$

where c_{RM} and C_{RM} are the measured and reference values of the analysed reference material, $u(c_{\mathrm{RM}})$ and $u(C_{\mathrm{RM}})$ the respective standard uncertainties and 3 the coverage factor to expand the standard uncertainty of the different, u_{d}, to approximately 99 % confidence level.

For instance, if the laboratory participates in a proficiency test reporting a nitrate mass concentration of (4.43 ± 0.82) mg L^{-1} for a reference value of (5.00 ± 0.20) mg L^{-1}, both using a coverage factor of 2 to expand the uncertainty to 95 % confidence level, the condition of Eq.(B1.17) applied in Eq.(B1.18) is valid making the described performance test an indication of uncertainty model adequacy.

$$\begin{gathered} |4.43 - 5.00| \leq 3\sqrt{(0.820/2)^2 + (0.20/2)^2} = \\ 0.57\ \mathrm{mg\ L^{-1}} \leq 1.27\ \mathrm{mg\ L^{-1}} \end{gathered} \qquad \text{(B1.18)}$$

B1.10. Limit of Quantification

According to Table B1.1, the c_{LOQ} is 0.33 mg L^{-1} ($c_{\mathrm{LOQ}} = 10 \cdot 0.033\ \mathrm{mg\ L^{-1}}$). Since c_{LOQ} is 0.33 mg L^{-1}, some laboratories might decide not to report below this concentration. However, the uncertainty calculations remain applicable between 0.2 mg L^{-1}and 0.3 mg L^{-1}.

B1.11. References

[B1.1] R. J. N. Bettencourt da Silva, "Spreadsheet for designing valid least-squares calibrations: A tutorial", *Talanta*, vol. 148, pp. 177-190, 2016.

[B1.2] R. J. N. Bettencourt da Silva, J. R. Santos and M. Filomena G. F. C. Camões, "Worst case uncertainty estimates for routine instrumental analysis", *Analyst*, vol. 127, pp. 957-963, 2002.

[B1.3] J. J. Delfino, "Stability of Nitrate in Unpreserved Potable Water Samples", *J. Am. Water Work Assoc.*, vol. 71, pp. 584-586, 1979.

[B1.4] ISO, Water quality - Sampling - Part 3: Preservation and handling of water samples (ISO5667-3), Geneva: ISO, 2018.

[B1.5] M. Roman, R. Dovi, R. Yoder, F. Dias and B. Warden, "Determination by ion chromatography and spectrophotometry of the effects of preservation on nitrite and nitrate", *J. Chromatogr. A*, vol. 546, pp. 341-346, 1991.

[B1.6] Council Directive 98/83/EC of 3 November 1998 on the quality of water intended for human consumption

[B1.7] Commission Directive (EU) 2015/1787 of 6 October 2015 amending Annexes II and III to Council Directive 98/83/EC on the quality of water intended for human consumption

[B1.8] JCGM, International Vocabulary of Metrology - Basic and general concepts and associated terms (JCGM 200), Sèvres: BIPM, 2012.
[B1.9] ILAC, G8:09/2019 Guidelines on Decision Rules and Statements of Conformity, Newton: ILAC, 2019.
[B1.10] A. Williams and B. Magnusson (eds.), Eurachem/CITAC Guide: Use of uncertainty information in compliance assessment (2nd ed.), Eurachem, 2021. ISBN 978-0-948926-38-9.
[B1.11] S. L. R. Ellison and A. Williams (eds.), Eurachem/CITAC guide: Quantifying Uncertainty in Analytical Measurement (3rd ed.), Eurachem, 2012. ISBN 978-0-948926-30-3.

B2: Determination of nitrate in drinking water by ion-chromatography – Uncertainty Evaluation II

Assessment of precision above and below typical nitrate values of analysed samples using Modelling 2 from Appendix A where the recovery is measured using two sets of spiked samples without native analyte when the recovery does not vary with the matrix

B2.1. Analytical method

See section B1.1.

B2.2. Stock solutions and sample conservation

See section B1.2.

B2.3. Measurement purpose

See section B1.3.

B2.4. Measurement requirements

See section B1.4.

B2.5. Analytical scope

See section B1.5.

B2.6. Specific feature of the example

This example describes a scenario of measurement performance assessment where (i) analysed samples are homogeneous, (ii) the intermediate measurement precision is quantified below, at and above typical values in samples (Modelling 2 from Appendix A) by pooling various precision estimates, and (iii) two sets of samples without native analyte were spiked daily below and at typical sample values producing reference materials equivalent to real samples.

B2.6.i. Uncertainty associated with precision

Two samples, (51/22) and (52/22), with nitrate mass concentrations approximately stable for eight days (refrigerated samples), and samples, without native analyte, spiked daily at (0.400 ± 0.018) mg L^{-1} (see section B1.6.i) and (0.2000 ± 0.0082) mg L^{-1} ($k = 2$ for 95 % confidence level) were analysed on different days to measure the intermediate precision below, at and above 0.400 mg L^{-1}.

B2.6.ii. Uncertainty associated with analyte recovery

The results from the analysis of spiked samples were used to measure the mean apparent analyte recovery. Mean recoveries estimated from two sets of data were compared to check their compatibility or metrological equivalence.

B2.6.iii. Mean recovery correction

See section B1.6.iii.

B2.7. Uncertainty calculations and quality control

B2.7.i. Precision uncertainty

Tables B1.1, B2.1, B2.2 and B2.3 present the results used to model the intermediate precision variation in the analytical interval.

At Interval I, [0.2 mg L^{-1}, 0.4 mg L^{-1}] excluding 0.4 mg L^{-1}, associated to sample solution quantified directly by IC, c_{IA}, (i.e. without or after dilution) the absolute intermediate precision standard deviation, $s_{\mathrm{I(P)}}\langle \mathrm{I} \rangle$, is obtained by pooling data from Tables B1.1, B2.1 and B2.2 using Eq.(B2.1). This equation is equivalent to Eq.(1), adapted to intermediate precision conditions.

$$s_{\mathrm{I(P)}}\langle \mathrm{I} \rangle = \sqrt{\frac{\sum_{i=1}^{N}(n_i - 1)s_{\mathrm{I}(i)}^2\langle \mathrm{I} \rangle}{\sum_{i=1}^{N}(n_i - 1)}} =$$

$$= \sqrt{\frac{(20-1)0.0328^2 + (6-1)0.0250^2 + (20-1)0.0245^2}{(20-1) + (6-1) + (20-1)}} = \tag{B2.1}$$

$$= 0.0285 \text{ mg L}^{-1}$$

All standard deviations pooled in Eq.(B2.1) are statistically equivalent at the 95 % confidence level.

For Interval II, [0.4 mg L^{-1}, 1.4 mg L^{-1}], the relative intermediate precision standard deviation of analysis summarised in Tables B.1.1 and B2.3 are pooled, giving a relative standard deviation, $s'_{\mathrm{I(P)}}\langle \mathrm{II} \rangle$, applicable at and above 0.4 mg L^{-1} to determinations of c_{IA} (see Eq.(2) adapted to intermediate precision conditions).

$$s'_{\mathrm{I(P)}}\langle \mathrm{II} \rangle = \sqrt{\frac{\sum_{i=1}^{N}(n_i - 1)s'^2_{\mathrm{I}(i)}\langle \mathrm{II} \rangle}{\sum_{i=1}^{N}(n_i - 1)}} =$$

$$= \sqrt{\frac{(20-1)8.43\ \%^2 + (6-1)4.11\ \%^2}{(20-1) + (6-1)}} = \tag{B2.2}$$

$$= 7.73\ \%$$

Table B2.1: Results from six measurements of nitrate in drinking water sample 51/22 stable, in refrigerated conditions, for at least eight days. The sample was analysed on different days. The mean, $\bar{c}_2$, standard deviation, $s_{\mathrm{I(2)}}$, and relative standard deviation, $s_{\mathrm{I(2)}}/\bar{c}_2$, are reported.

Date	Sample	Concentration, c_2/mg L^{-1}
05/01/2022	51/22	0.288
06/01/2022		0.345
07/01/2022		0.315
10/01/2022		0.272
11/01/2022		0.303
12/01/2022		0.297
	$\bar{c}_2$	0.3033 mg L^{-1}
	$s_{\mathrm{I(2)}}$	0.0250 mg L^{-1}
	$s_{\mathrm{I(2)}}/\bar{c}_2$	8.24 %
	n_2	6

Both relative standard deviations pooled in Eq.(B2.2) are statistically equivalent at the 95 % confidence level.

Therefore, for Interval I, $s_{\mathrm{I(P)}}\langle \mathrm{I} \rangle = 0.0285\ \mathrm{mg\ L^{-1}}$ and for Interval II, $s'_{\mathrm{I(P)}}\langle \mathrm{II} \rangle = 7.73\ \%$; in this case using a transition point of $0.4\ \mathrm{mg\ L^{-1}}$ equivalent to the typical values in samples. Alternatively, a transition point, c_{T}, of $0.37\ \mathrm{mg\ L^{-1}}$ can be considered ($c_{\mathrm{T}} = 0.0285/0.0773$) assuming $s_{\mathrm{I\langle L\rangle}} = 0.0285\ \mathrm{mg\ L^{-1}}$ and $s'_{\mathrm{I\langle H\rangle}} = 0.0773$. In that case, for Interval I ([0.2 mg L^{-1}, 0.37 mg L^{-1}] excluding 0.37 mg L^{-1}), $s_{\mathrm{I(P)}}\langle \mathrm{I} \rangle = 0.0285\ \mathrm{mg\ L^{-1}}$ and in Interval II, [0.37 mg L^{-1}, 1.4 mg L^{-1}], $s'_{\mathrm{I\langle H\rangle}} = 0.0773$.

Table B2.2: Results from the measurement of nitrate in twenty, n_3, drinking water samples without detectable nitrate concentrations spiked at (0.2000 ± 0.0082) mg L^{-1} (coverage factor, k, equal to 2 for approximately 95 % confidence level). Samples were analysed on different days. The mean, $\bar{c}_3$, standard deviation, $s_{\mathrm{I(3)}}$, and relative standard deviation, $s_{\mathrm{I(3)}}/\bar{c}_3$, are reported.

Date	Sample	Concentration, c_3/ mg L^{-1}
05/01/2022	50/22	0.194
06/01/2022	53/22	0.193
07/01/2022	107/22	0.227
10/01/2022	117/22	0.234
11/01/2022	124/22	0.197
(…)	(…)	(…)
28/01/2022	270/22	0.197
31/01/2022	279/22	0.197
03/02/2022	285/22	0.242
04/02/2022	286/22	0.218
07/02/2022	289/22	0.179
	$\bar{c}_3$	0.204 mg L^{-1}
	$s_{\mathrm{I(3)}}$	0.0245 mg L^{-1}
	$s_{\mathrm{I(3)}}/\bar{c}_3$	12.01 %
	n_3	20

Table B2.3: Results from six measurements of nitrate in drinking water sample 52/22 stable, in refrigerated conditions, for at least eight days. The sample was analysed on different days. The mean, $\bar{c}_4$, standard deviation, $s_{\mathrm{I(4)}}$, and relative standard deviation, $s_{\mathrm{I(4)}}/\bar{c}_4$, are reported.

Date	Sample	Concentration/ mg L^{-1}
05/01/2022	52/22	0.773
06/01/2022		0.732
07/01/2022		0.772
10/01/2022		0.817
11/01/2022		0.803
12/01/2022		0.812
	$\bar{c}_4$	0.7848 mg L^{-1}
	$s_{\mathrm{I(4)}}$	0.0322 mg L^{-1}
	$s_{\mathrm{I(4)}}/\bar{c}_4$	4.11 %
	n_4	6

The *Quality control* described in section B1.7.i. is applicable to this validation scenario. Figure B2.1 illustrates the models developed to describe the variation of intermediate precision with the measured nitrate mass concentration.

B2.7.ii. Recovery uncertainty

From the analysis of samples without native analyte spiked at 0.4 mg L^{-1} of nitrate, $\bar{R}_1$ = 97.2 % and $u_{\bar{R}1}$ = 2.85 %; this $\bar{R}_1$ is not significantly different from 100 % (see section B1.7.ii).

If the same relations are used to study the results from the analysis of samples spiked at 0.2 mg L^{-1}, then $\bar{R}_2 = 0.204/0.2 = 102.0\ \%$ and a $u_{\bar{R}2} = 1.02((0.0245/(0.204 \cdot (20)^{1/2}))^2 + (0.0082/2/0.2)^2)^{1/2} = 3.45\ \%$. The $\bar{R}_2 = 102.0\ \%$ is not significantly different from 100 %.

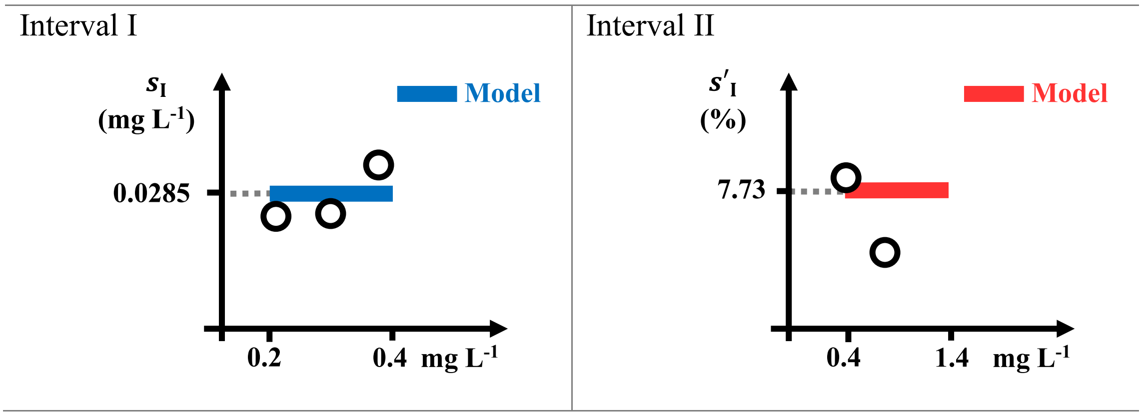

Figure B2.1. Model of variation of intermediate precision with the measured nitrate mass concentration by IC. The model is applicable to direct measurements (i.e., after required dilutions) and was developed from results of the analysis of two samples and two spiked samples sets.

Both mean recoveries are also equivalent since the following condition for compatibility of results is valid:

$$
\begin{gathered}
|\bar{R}_1 - \bar{R}_2| \leq 3\sqrt{u_{\bar{R}1}^2 + u_{\bar{R}2}^2} \Leftrightarrow \\
\Leftrightarrow |0.972 - 1.02| \leq 3\sqrt{0.0285^2 + 0.0345^2} \Leftrightarrow \\
\Leftrightarrow 0.048 \leq 0.134
\end{gathered}
\qquad \text{(B2.3)}
$$

Therefore, the mean recoveries are used to estimate an overall mean recovery, $\bar{\bar{R}}_{12}$, and respective standard uncertainty, $u_{\bar{\bar{R}}12}$, following Eq.(B2.4) and (B2.5) which are equivalent to the use of Eq.(8) and (9).

$$
\bar{\bar{R}} = \frac{0.972 + 1.02}{2} = 99.6\ \% \qquad \text{(B2.4)}
$$

$$u_{\bar{\bar{R}}} = \frac{\sqrt{\sum_{i=1}^{M}\left\{\left(\frac{\bar{c}_i}{C_i}\right)^2\left[\left(\frac{s_{\mathrm{I}(i)}}{\bar{c}_i\sqrt{n_i}}\right)^2+\left(\frac{u(C_i)}{C_i}\right)^2\right]\right\}}}{N} = \frac{\sqrt{\begin{array}{c}\left\{\left(\frac{0.389}{0.4}\right)^2\left[\left(\frac{0.0328}{0.389\sqrt{20}}\right)^2+\left(\frac{0.009}{0.4}\right)^2\right]\right\}+\\ +\left\{\left(\frac{0.204}{0.2}\right)^2\left[\left(\frac{0.0245}{0.204\sqrt{20}}\right)^2+\left(\frac{0.0041}{0.2}\right)^2\right]\right\}\end{array}}}{2} = =2.24\ \% \qquad \text{(B2.5)}$$

The overall mean recovery (i.e. 99.6 %) is still equivalent to 100% given their uncertainty (i.e. 2.24 %).

Quality control

A quality control equivalent to the one described in section B1.7.ii. for daily recovery values can be defined for samples spiked at 0.2 mg L^{-1} or both spiked levels using $\bar{\bar{R}}$ and $u_{\bar{\bar{R}}}$ associated with approximately $\left((20-1)+(20-1)\right)=38$ degrees of freedom.

B2.7.iii. Additional uncertainty components

See section B1.7.iii.

B2.7.iv. Uncertainty combination and expansion

The uncertainty components are combined as described in Eq.(B2.6) and (B2.7) for Interval I and II, respectively.

$$\begin{aligned} U\langle \mathrm{I}\rangle &= 2\sqrt{s^2_{\mathrm{I(p)}}\langle \mathrm{I}\rangle + (c_\mathrm{s}\cdot u'_{\bar{\bar{R}}})^2 + 2(c_\mathrm{s}\cdot u'_{\mathrm{DSS}})^2 + (c_\mathrm{s}\cdot u'_F)^2} = \\ &= 2\sqrt{s^2_{\mathrm{I(p)}}\langle \mathrm{I}\rangle + c_\mathrm{s}^2\left(u'^2_{\bar{\bar{R}}} + 2u'^2_{\mathrm{DSS}} + u'^2_F\right)} = \\ &=2\sqrt{0.0285^2 + c_\mathrm{S}^2((0.0224/1)^2 + 2(0.0095)^2 + (0.022)^2)} = \\ &=2\sqrt{0.0285^2 + c_\mathrm{s}^2(0.001166)} \end{aligned} \qquad \text{(B2.6)}$$

$$\begin{aligned} U\langle \mathrm{II}\rangle &= 2c_\mathrm{s}\sqrt{s'^2_{\mathrm{I(P)}}\langle \mathrm{II}\rangle + u'^2_{\bar{\bar{R}}} + 2u'^2_{\mathrm{DSS}} + u'^2_F} = \\ &=2c_\mathrm{S}\sqrt{0.0773^2 + (0.0224/1)^2 + 2(0.0095)^2 + (0.022)^2} = \\ &= 2c_\mathrm{S}0.0845 = c_\mathrm{S}0.169 \end{aligned} \qquad \text{(B2.7)}$$

where $U\langle \mathrm{I}\rangle$ and $U\langle \mathrm{II}\rangle$ are expanded uncertainties for a 95 % confidence level applicable to c_{IA} below and above 0.4 mg L^{-1}, respectively, and c_s is the measured concentration of the analysed sample ($c_\mathrm{s} = c_{\mathrm{IA}}\cdot F$).

For a sample with a measured mass concentration of 50 mg L^{-1}, $U\langle \mathrm{II}\rangle = 50\cdot 0.169 = 8.45$ mg L^{-1}. This measurement uncertainty is still larger than the maximum permissible value of 7.5 mg L^{-1} [B1.7].

B2.7.v. Uncertainty contributions

General equations B1.8 to B1.15 can be used to estimate the percentage contribution of the various uncertainty components to the measurement uncertainty.

B2.8. Uncertainty reduction

To ensure measurements at 57.5 mg L^{-1} have adequately low uncertainty, the sample must be analysed on two days ($p = 2$), where on each day, a single analysis ($p = 1$) is performed. Equation B2.8 presents the calculations performed. In this case, the diluted sample solutions must have a mass concentration between 0.4 mg L^{-1} and 1.4 mg L^{-1}.

$$U\langle\mathrm{II}\rangle = 2 \cdot 57.5\sqrt{\begin{array}{c}\left[\dfrac{s'^2_{\mathrm{I(P)}}\langle\mathrm{II}\rangle}{p}\right] + (0.0224/1)^2 + \\ +2(0.0095)^2 + (0.022)^2\end{array}} = \\ = 2 \cdot 57.5\sqrt{\begin{array}{c}\left[\dfrac{0.0773^2}{2}\right] + (0.0224/1)^2 + \\ +2(0.0095)^2 + (0.022)^2\end{array}} = \\ = 7.41\ \mathrm{mg\ L^{-1}} \qquad \text{(B2.8)}$$

If it is decided to analyse the sample on two different days ($p = 2$), but on each day, the sample is analysed three times under repeatability conditions ($n = 3$), the mean of six values is given with an expanded uncertainty calculated in Eq.(B2.9). (see Eq.(5))

$$U\langle\mathrm{II}\rangle = 2 \cdot 57.5\sqrt{\begin{array}{c}\left[\dfrac{s'^2_{\mathrm{I(P)}}\langle\mathrm{II}\rangle}{p} + \dfrac{s'^2_{\mathrm{r}}\langle\mathrm{II}\rangle(1-n)}{np}\right] + (0.0224/1)^2 + \\ +2(0.0095)^2 + (0.022)^2\end{array}} = \\ = 2 \cdot 57.5\sqrt{\begin{array}{c}\left[\dfrac{0.0773^2}{2} + \dfrac{0.0340^2(1-3)}{2 \cdot 3}\right] + (0.0224/1)^2 + \\ +2(0.0095)^2 + (0.022)^2\end{array}} = \\ = 7.06\ \mathrm{mg\ L^{-1}} \qquad \text{(B2.9)}$$

The determination of $s'_{\mathrm{r}}\langle\mathrm{II}\rangle = 0.0136/0.4 = 0.034$ is mentioned in Example B1 (section B1.7.i.; Quality Control). The replicate analysis of drinking water samples allows the analysis in 48h and is only necessary for determinations of nitrate mass concentrations between 44.5 mg L^{-1} and 57.5 mg L^{-1}. Therefore, the described method, including the relevant evaluation of measurement uncertainty of results, is fit for monitoring drinking waters following European Union legislation.

B2.9. Uncertainty model assumptions and limitations

Equivalent to the description of section B1.9 with an addendum related to information on analyte recovery close to 0.2 mg L^{-1} that has improved the representativeness of the uncertainty evaluation.

B2.10. Limit of Quantification

See section B1.10.

B2.11. References

See section B1.11.

B3: Determination of nitrate in drinking water by ion-chromatography – Uncertainty Evaluation III

Assessment of precision at four analyte intervals by using data from various samples using Modelling 1 and 4 from Appendix A where recovery is measured using spiked samples without native analyte when the recovery does not vary with the matrix

B3.1. Analytical method

See section B1.1.

B3.2. Stock solutions and sample conservation

See section B1.2.

B3.3. Measurement purpose

See section B1.3.

B3.4. Measurement requirements

See section B1.4.

B3.5. Analytical scope

See section B1.5.

B3.6. Specific feature of the example

This example describes a scenario of measurement performance assessment where (i) analysed samples are homogeneous, (ii) the intermediate measurement precision is quantified from a linear regression between observed intermediate precision standard deviation and measured concentration using the same data considered in Example B2, (iii) and analyte recovery was assessed as described in Example B2. The model of the variation of intermediate precision with measured nitrate concentration was defined from the same data used in Example B2 but following different modelling criteria.

B3.6.i. Uncertainty associated with precision
See section B2.6.i.

B3.6.ii. Uncertainty associated with analyte recovery
See section B2.6.ii.

B3.6.iii. Recovery correction option
See section B1.6.iii.

B3.7. Uncertainty calculations and quality control

B3.7.i. Precision uncertainty
The same tables B1.1, B2.1, B2.2 and B2.3 used in Example B2, are used to produce an alternative more complex model of the variation of intermediate precision with the concentration measured by IC. Given the summary of collected intermediate precision data in Table B3.1, since the studied concentration levels are similar, the relation between s_I vs. c from 0.20 mg/L to 0.78 mg/L was considered to be linear ($s_\mathrm{I} = 0.0233 + 0.0128c$, with a Pearson’s correlation coefficient of 0.7266). This relation suggests an approximately constant s_I in the studied interval studied (Modelling 4 of Appendix A). Above 0.78 mg/L, a worst case $s'_\mathrm{I} = 4.11\ \%$ is considered (Modelling 1 of Appendix A).

Table B3.1: Summary of intermediate precision estimated from the replicate analysis of nitrate in drinking water samples (see Table B1.1, B2.1, B2.2 and B2.3).

Mean value, *c*/ mg L^{-1}	s_I/ mg L^{-1}	s'_I	*n*
0.3890	0.0328	8.43%	20
0.3033	0.0250	8.24%	6
0.2040	0.0245	12.01%	20
0.7848	0.0322	4.11%	6

Figure B3.1 illustrates the models developed to describe the variation of intermediate precision with the measured nitrate mass concentration.

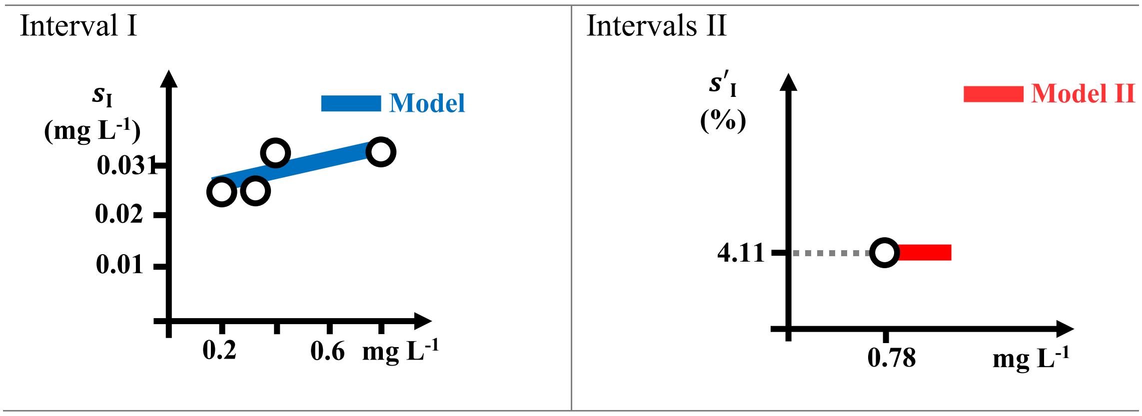

Figure B3.1. Model of the variation of intermediate precision with the measured nitrate mass concentration by IC. The model is applicable to direct measurements (i.e., after required dilutions) and was developed from results of the analysis of two samples and two spiked samples sets.

B3.7.ii. Recovery uncertainty
See section B2.7.ii.

B3.7.iii. Additional uncertainty components
See section B1.7.iii.

B3.7.iv. Uncertainty combination and expansion:
The uncertainty components are combined as described in Eq.(B3.1) and (B3.2) for Intervals I and II, respectively.

$$U\langle \mathrm{I}\rangle = 2\sqrt{F^2(0.0233+0.0128c_\mathrm{IA})^2 + c_\mathrm{S}^2\left(u'^2_{\bar{\bar{R}}} + 2u'^2_\mathrm{DSS} + u'^2_F\right)} = \\ = \\ 2\sqrt{F^2(0.0233+0.0128c_\mathrm{IA})^2 + c_\mathrm{S}^2[(0.0224/1)^2 + 2(0.0095)^2 + (0.022)^2} \\ = 2\sqrt{F^2(0.0233+0.0128c_\mathrm{IA})^2 + c_\mathrm{S}^2[0.00117]} \qquad \text{(B3.1)}$$

$$U\langle \mathrm{II}\rangle = 2.57c_\mathrm{S}\sqrt{s'^2_\mathrm{I}\langle \mathrm{II}\rangle + u'^2_{\bar{\bar{R}}} + 2u'^2_\mathrm{DSS} + u'^2_F} = \\ =2.57c_\mathrm{S}\sqrt{0.0411^2 + (0.0224/1)^2 + 2(0.0095)^2 + (0.022)^2} = \\ = 2.57c_\mathrm{S}0.0534 = c_\mathrm{S}0.137 \qquad \text{(B3.2)}$$

where $U\langle\mathrm{I}\rangle$ and $U\langle\mathrm{II}\rangle$ are expanded uncertainties for a 95 % confidence level applicable to c_{IA} determined between 0.20 mg L^{-1} and 0.78 mg L^{-1}, and above 0.78 mg L^{-1}, respectively, and c_{s} is the measured concentration of the analysed sample ($c_{\mathrm{s}} = c_{\mathrm{IA}} \cdot F$; F is the dilution factor). For Interval II, since the larger uncertainty component has only 5 degrees of freedom ($s'_{\mathrm{I(P)}}\langle\mathrm{II}\rangle$ was estimated from six replicates), the more conservative approach of expanding the standard uncertainty by using t value of the Student's t distribution for a 95 % confidence level and $(6-1)$ degrees of freedom (i.e., 2.57) was followed.

For a sample with a measured mass concentration, c_{s}, of 57.5 mg L^{-1}, determined from the analysis of a diluted sample solution with a $c_{\mathrm{IA}} = 0.575\ \mathrm{mg\,L^{-1}}$ after a 100 times dilution $(F = 100)$, $U\langle\mathrm{I}\rangle = 2\sqrt{100^2(0.0233 + 0.0128 \cdot 0.575)^2 + 57.5^2[0.00117]} = 7.28\ \mathrm{mg\,L^{-1}}$. This measurement uncertainty is lower than the maximum permissible value of 7.5 mg L^{-1} (see section B1.4).

B3.7.v. Uncertainty contributions

General equations B1.8 to B1.15 can be used to estimate the percentage contribution of the various uncertainty components to the measurement uncertainty.

B3.8. Uncertainty model assumptions and limitations

See section B2.9.

B3.9. Limit of Quantification

See section B1.10.

B3.10. References

See section B1.11.

Conclusions from examples B1, B2 and B3

Examples B1, B2 and B3 illustrate how the available data on measurement performance and how modelling criteria affect the evaluated measurement uncertainty. All approaches presented are valid and require being checked on a routine basis through analysis of other reference materials to confirm evaluation validity.

B4: Determination of total As in sediments – Uncertainty evaluation I

Assessment of precision affected by sample heterogeneity using Modelling 1 from Appendix A and recovery estimated from the analysis of a certified reference material

B4.1. Analytical method

The determination of As in a marine sediment sample received in the laboratory by hydride generation atomic absorption spectrometry (HGAAS) after microwave assisted acid digestion is performed as summarised below following a procedure described by OSPAR [B4.1]:

(1) Sediment is sieved, lyophilised and milled; (2) a sediment portion of (0.500 ± 0.050) g is weighted; (3) the portion is acid digested with *aqua regia* and hydrofluoric acid in a closed vessel irradiated with microwaves; (4) after sample solution cooling, the hydrofluoric acid excess is neutralised with boric acid; (5) the sample solution volume is adjusted to 50 mL and diluted 50 times more; (6) the As in the diluted sample solution is quantified by HGAAS after calibration below 5 µg L^{-1} where the arsenic of calibrators and sample solutions is reduced to As^{3+} before atomisation.

More details about the method used are available elsewhere [B4.2].

B4.2. Stock solutions and sample conservation

Sediment samples are preserved frozen below -20 ºC for 6 months and stock solutions and reference materials stored according to manufacturer instructions.

B4.3. Measurement purpose

The analytical method is used to assign sediments to the respective contamination level class according to the Portuguese law, Portaria nº 1450/2007 [B4.3], to decide whether sediments can be used for beach nourishment or dredging rules. Dredging rules are defined to protect the environment from the dragged area and the places where sediments are deposited. Regarding total As levels, five classes are defined: Class 1 (< 20 mg kg^{-1}), Class 2 (20 mg kg^{-1} to 50 mg kg^{-1}), Class 3 (50 mg kg^{-1} to 100 mg kg^{-1}), Class 4 (100 mg kg^{-1} to 500 mg kg^{-1}) and Class 5 (> 500 mg kg^{-1}).

There is no European Union law for the management of sediments dragging.

B4.4. Measurement requirements

The Portaria nº 1450/2007 [B4.3] does not specify measurement quality requirements in particular no target (maximum admissible) uncertainty is given. However, according to the Eurachem/CITAC guide on setting target measurement uncertainty [B4.4], the class interval range can be used to set a target expanded uncertainty, U^{tg} : $U^{\text{tg}} = (Q^{\max} - Q^{\min})/8$, where $Q^{\max}$ and $Q^{\min}$ are the maximum and minimum interval limits. Following this criteria, for classes 2, 3 and 4 the U^{tg} should be 3.75 mg kg^{-1}, 6.25 mg kg^{-1} and 50 mg kg^{-1}, respectively. For classes 1 and 5 a U^{tg} of 3.75 mg kg^{-1} and 50 mg kg^{-1} is considered above 16.25 mg kg^{-1} and below 550 mg kg^{-1} where distinction between two classes is more difficult. The decision rule of conformity assessments should be set by the national regulator for these type of assessments.

B4.5. Analytical scope

B4.5.i. Matrices

Marine or river sediments.

B4.5.ii. Analytical interval

The atomic spectrometer is calibrated using a blank signal and signals from quantitative calibrators between 0.5 µg L^{-1} and 5 µg L^{-1}. If the 50 mL digested sample solution is not diluted, mass fractions between 0.05 mg kg^{-1} and 0.5 mg kg^{-1} can be quantified. The defined and validated limit of quantification, w_{LOQ}, of the analytical method is 0.05 mg kg^{-1}. If the sample solution is diluted x times with class A volumetric pipettes and flasks with volumes not lower than 0.5 mL and 5 mL, respectively, the limits of this interval can be increased x times with dilutions with negligible uncertainty given the contribution of other components to the overall measurement uncertainty. The validated lower and upper mass fraction limits depend on As mass fractions of sediments considered in method validation. In this case, the validated mass fraction interval is between 1.3 mg kg^{-1} and 25 mg kg^{-1} as it will be discussed below.

B4.6. Specific feature of the example

This example describes a scenario where the heterogeneity of analysed samples can be a relevant uncertainty component. The measurement precision is evaluated from the intermediate precision of the analysis of real samples that includes samples heterogeneity, and measurement trueness is assessed from the analysis of one Certified Reference Material, CRM. The limited number of studied As mass fractions and reference materials considered in the trueness assessment should be kept in mind when using the developed uncertainty model. Limited scope uncertainty evaluations should be overcome every time considered relevant by considering the analysis of additional reference materials in the uncertainty evaluation.

B4.6.i. Precision uncertainty

Intermediate measurement precision was estimated from the analysis of three real sediment samples (A, B and C) where sample A was analysed in duplicate under repeatability conditions.

B4.6.ii. Mean analyte recovery uncertainty

Analyte recovery was estimated from the analysis of the CRM WQB-1 from lake Ontario sediments [B4.5]. This reference material was analysed ten times on ten different days.

B4.6.iii. Mean recovery correction

If the estimated mean analyte recovery is significantly different from 100 % and the deviation not attributed to the lack of linearity of atomic spectrometer response or digestion efficiency saturation, the results should be corrected for observed mean recovery. Nevertheless, the matrix effect observed using the CRM can be different from that affecting the sample analysed due to matrix mismatch in terms of analytical performance. However, the option to evaluate the uncertainty from the analysis of a single reference material assumes method performance is equivalent for the reference material and analysed unknown samples.

B4.7. Uncertainty calculations

B4.7.i. Precision uncertainty

Measurement precision was estimated from the analysis of real marine sediment samples A, B and C with estimated mean mass fractions of 15.8 mg kg^{-1}, 14.6 mg kg^{-1} and 6.5 mg kg^{-1}, respectively. Sample A was analysed twice, under repeatability conditions, but the intermediate precision from the first and second replicate (A1 and A2) was considered independently. Table B4.1 presents the estimated intermediate precision from the analysis of samples A (replicate A1 and A2), B and C [B4.2].

Table B4.1. Estimated mean mass fractions, $\bar{w}$, intermediate precision absolute, s_I, and relative, s'_I, standard deviations and number of replicate analysis, *n*, of real marine sediment samples A, B and C. Duplicates of sample A (A1 and A2) were analysed under repeatability conditions.

Sample	$\bar{w}$/ mg kg^{-1}	s_I/ mg kg^{-1}	s'_I	n
A1	15.693	0.460	2.93 %	10
A2	15.88	1.97	12.4 %	10
B	14.599	0.797	5.46 %	10
C	6.461	0.312	4.83 %	9

Note: $\bar{w}$ and s_I are reported with the same decimal places that allow a s_I with three significant figures and s'_I with three significant figures. This number of figures is adequate for calculation precision.

Analysed sediments A, B and C have equivalent total As mass fractions well above the LOQ (w_LOQ) where s'_I tends to become approximately constant. The s'_I from the duplicate analysis of Sample A (i.e. A2) is significantly different from that observed for the other samples. Although this difference can be attributed to the heterogeneity of sediment samples, it was considered an outlier in precision evaluation. Equation B4.1 presents the pooling of estimated s'_I for samples A1, B and C.

$$s'_{\mathrm{I(P)}}\langle\mathrm{II}\rangle = \sqrt{\frac{\sum_{i=1}^{N}(n_i-1)s'^2_{\mathrm{I}(i)}\langle\mathrm{II}\rangle}{\sum_{i=1}^{N}(n_i-1)}} =$$

$$= \sqrt{\frac{(10-1)2.93\ \%^2 + (10-1)5.46\ \%^2 + (9-1)4.83\ \%^2}{(10-1)+(10-1)+(9-1)}} = \tag{B4.1}$$

$$= 4.52\ \%$$

Therefore, the following model of precision variation with total As mass fraction was considered.

Interval I (between 1.3 mg kg^{-1} (one fifth of 6.46 mg kg^{-1}) and 6.46 mg kg^{-1}) [1.3 mg kg^{-1}; 6.46 mg kg^{-1}]: Constant $s_\mathrm{I}\langle\mathrm{I}\rangle = 0.292\ \mathrm{mg\,kg^{-1}}$ ($4.52\ \% \cdot 6.46\ \mathrm{mg\,kg^{-1}} = 0.292\ \mathrm{mg\,kg^{-1}}$). It is considered that using $s_\mathrm{I} = 0.292\ \mathrm{mg\,kg^{-1}}$ below one fifth of 6.46 mg kg^{-1} (i.e. 1.3 mg kg^{-1}) would be too pessimistic.

Interval II [6.46 mg kg^{-1}; 25 mg kg^{-1}]: Constant s'_I of 4.52 %, $s'_{\mathrm{I(P)}}\langle\mathrm{II}\rangle$, estimated from Eq.(B4.1). 25 mg kg^{-1} is 20 %* above the larger mass fraction analysed for method validation through the analysis of the CRM (see section B4.7.ii).

B4.7.ii. Mean analyte recovery uncertainty

Mean analyte recovery, $\bar{R}$, was estimated from the replicate analysis of the CRM WQB-1 [B4.5] with a reference total As value of (23.00 ± 1.84) mg kg^{-1} ($W_\mathrm{RM} \pm U(W_\mathrm{RM})$) for a coverage factor, *k*, of 2 associated with a 95 % confidence level. The reference value is

---

* Instead of a 20 % tolerance, it can be considered precision from the analysis of the CRM to define the mass fraction range.

defined for the analysis of portions not lower than 0.2 g. This reference material was analysed 10, n, times on different days. The results had a mean of 21.040 mg kg^{-1}, $\bar{w}_{\mathrm{RM}}$, and a standard deviation of 0.779 mg kg^{-1}, $s_{\mathrm{I}}(w_{\mathrm{RM}})$. $\bar{R}$ is 0.9148 (i.e., 21.040/23.00) and its standard uncertainty, $u_{\bar{R}}$, is 0.0381 evaluated from Eq.(B4.2) following the algorithm of Eq.(9) for N equal to 1.

$$u_{\bar{R}} = \sqrt{\left(\frac{\bar{w}_{\mathrm{RM}}}{W_{\mathrm{RM}}}\right)^2 \left[\left(\frac{s_{\mathrm{I}}(w_{\mathrm{RM}})}{\bar{w}_{\mathrm{RM}}\sqrt{n}}\right)^2 + \left(\frac{U(W_{\mathrm{RM}})/k}{W_{\mathrm{RM}}}\right)^2\right]} = \sqrt{\left(\frac{21.040}{23.00}\right)^2 \left[\left(\frac{0.779}{21.040\sqrt{10}}\right)^2 + \left(\frac{1.84/2}{23.00}\right)^2\right]} = = 0.0381 \qquad \text{(B4.2)}$$

The observed mean recovery is not significantly different from 100 % as proven from Eq.(B4.3). Results for test samples do not therefore require correction for observed recovery.

$$\frac{|1-\bar{R}|}{u_{\bar{R}}} \leq t_{\nu}^{0.95} \Leftrightarrow \Leftrightarrow \frac{|1-0.9148|}{0.0381} \leq t_{(10-1)}^{0.95} \Leftrightarrow \Leftrightarrow 2.24 \leq 2.26 \qquad \text{(B4.3)}$$

B4.7.iii. Additional uncertainty components

No significant additional uncertainty components were identified. Since the same stock solution of As was used to prepare calibrators for the replicate analysis of samples A, B and C, and for the analysis of the CRM, the stock solution uncertainty should be considered as an additional uncertainty component. However, given the uncertainty from complex sample preparation and sediment heterogeneity, this uncertainty component is negligible. The sample solution dilution is also not specified in the measurement uncertainty model since the dilution procedure guarantees dilution with negligible uncertainty. (Procedure: solution aliquot and final volume not lower than 0.5 mL and 5 mL, respectively, measured with class A volumetric glassware).

B4.7.iv. Uncertainty combination and expansion:

The expanded uncertainty of estimated sample mass fraction, w, for approximately 95 %, confidence interval, varies with the two intervals defined for the analytical method scope. Since $\bar{R}$ is equivalent to 1 it is assumed that absolute, $u_{\bar{R}}$, and relative, $u'_{\bar{R}}$, mean recovery standard uncertainties are equivalent ($u'_{\bar{R}} = u_{\bar{R}}/1$) (see section 8.2.6.2).

Interval I [1.3 mg kg^{-1}; 6.46 mg kg^{-1}]:

$$U\langle \mathrm{I}\rangle = 2\sqrt{s_{\mathrm{I}}^2\langle \mathrm{I}\rangle + \left(w \cdot \frac{u_{\bar{R}}}{1}\right)^2} = = 2\sqrt{0.292^2 + (w \cdot 0.0381)^2} \qquad \text{(B4.4)}$$

Interval II [6.46 mg kg^{-1}; 25 mg kg^{-1}]:

$$U\langle \mathrm{II}\rangle = 2w\sqrt{s'^2_{\mathrm{I}}\langle \mathrm{II}\rangle + \left(\frac{u_{\bar{R}}}{1}\right)^2} = \\ = 2w\sqrt{0.0452^2 + 0.0381^2} = 0.118w \qquad (B4.5)$$

For instance, for a w equal to 16 mg kg^{-1} from Interval II, U is 1.89 mg kg^{-1} with the result being reported as (16.0 ± 1.9) mg kg^{-1} for a coverage factor of $k = 2$ and approximately 95 % confidence level.

Figure B4.1 presents the model of the variation of the expanded uncertainty, U, and defined target expanded uncertainty, U^{tg}, in the validated mass fraction interval. This figures shows that measurements between 1.3 mg kg^{-1} and 25 mg kg^{-1} have fit for purpose uncertainty.

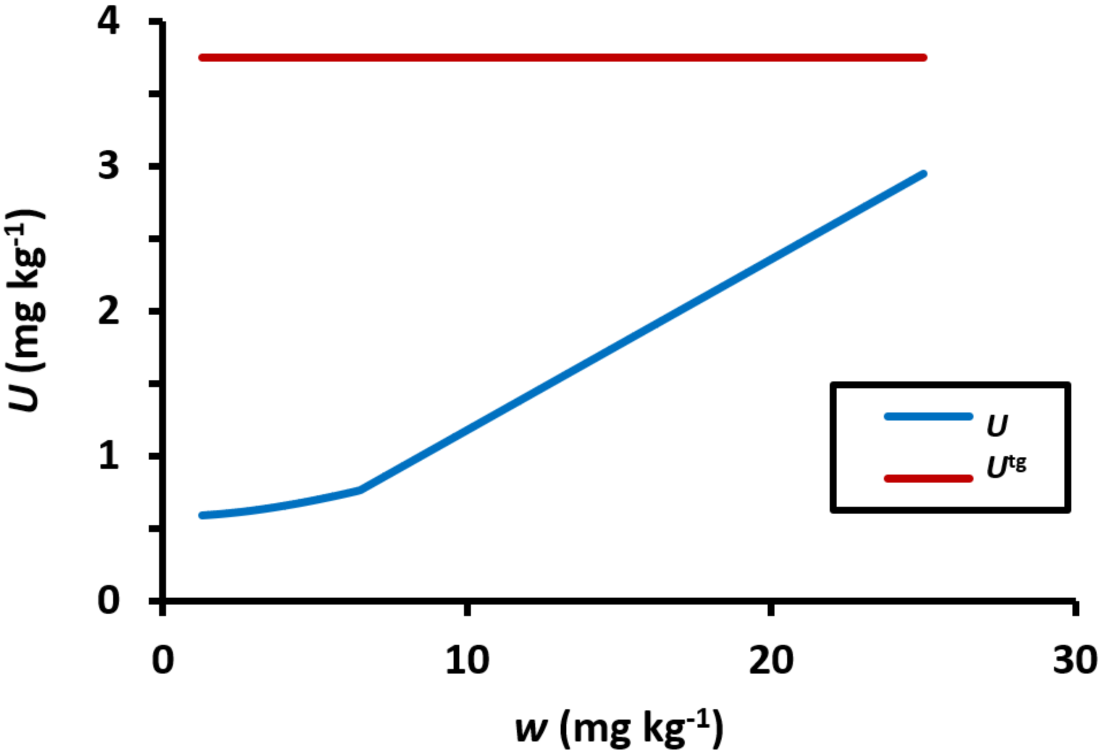

Figure B4.1. Model of measurement, U, and target, U^{tg}, expanded uncertainties, for 95 % confidence level, for the measured total As mass fraction, w, defined from the intermediate precision estimated from the analysis of three real marine sediments and analyte recovery determined from the analysis of a certified reference material.

B4.7.v. Uncertainty contributions

The same equations presented previously can be used to quantify the percentage contribution of the uncertainty components that vary with the mass fraction value.

B4.8. Uncertainty model assumptions and limitations

The described measurement uncertainty model is applicable to the determination of total As, between 1.3 mg kg^{-1} and 25 mg kg^{-1}, for sediments not more heterogenous than analysed sediments A, B and C and affected by matrix effects equivalent to that observed from the analysis of the CRM WQB-1. The validity of the described measurement model can be checked through the analysis of other reference materials as described in section B1.9.

B4.9. References

[B4.1] OSPAR, JAMP Guidelines for Monitoring Contaminants in Sediments, (Agreement Ref. No. 2002-16), London: OSPAR, 2002.

[B4.2] V. Morgado, C. Palma and R. J. N. Bettencourt da Silva, “Monte Carlo bottom-up evaluation of the uncertainty of complex sample preparation: Elemental determination in sediments”, *Anal. Chim. Acta*, vol. 1175, 338732, 2021.

[B4.3] Portaria 1450/2007 de 12 de Novembro de 2007 - Regras do regime de utilização dos recursos hídricos.

[B4.4] R. Bettencourt da Silva and A. Williams (eds.), Eurachem/CITAC Guide: Setting and Using Target Uncertainty in Chemical Measurement, Eurachem, 2015.

[B4.5] NWRI, Certificate of Analysis - WQB-1 - A Lake Ontario Sediment for Trace Elements, Burlington: NWRI, 2001.

B5: Determination of total As in sediments – Uncertainty evaluation II

Assessment of precision affected by sample heterogeneity using Modelling 1 from Appendix A and recovery estimated from the analysis of a certified reference material and spiked samples with native analyte where recovery varies with the matrix

B5.1. Analytical method

See section B4.1.

B5.2. Stock solutions and sample conservation

See section B4.2.

B5.3. Measurement purpose

See section B4.3.

B5.4. Measurement requirements

See section B4.4.

B5.5. Analytical scope

See section B4.5.

B5.6. Specific feature of the example

This example describes a scenario where the heterogeneity of analysed samples can be a relevant uncertainty component. The measurement precision is evaluated from the intermediate precision of the analysis of real samples that includes samples heterogeneity, and measurement trueness is assessed from the analysis of one Certified Reference Material, CRM, and two spiked samples with native analyte.

B5.6.i. Precision uncertainty

Evaluated from the same data as in Example B4 (See subsection B4.6.i).

B5.6.ii. Mean analyte recovery uncertainty

Analyte recovery was estimated from the analysis of the CRM WQB-1 from lake Ontario sediments [B4.5] and two spiked samples with native analyte. These reference materials were analysed ten times on ten different days.

B5.6.iii. Mean recovery correction

See subsection B4.6.iii. In this case, data from analyte recovery on spiked samples improves the assessment of measurement trueness by taking into account additional matrices realities.

B5.7. Uncertainty calculations

B5.7.i. Precision uncertainty

See subsection B4.7.i.

B5.7.ii. Mean analyte recovery uncertainty

The analyte recovery was estimated from the analysis of the CRM WQ-1 described in subsection B4.7.ii and the analysis of real sediment samples A and C before and after spiking with 3.761 µg of As with a relative standard uncertainty of 0.92 %. Regarding the analysis of sample A before spiking, it is considered means A1&2 of duplicate determinations (A1 and A2) performed under repeatability conditions (i.e., the standard deviation of daily means, $s_{\mathrm{I}}(w_{\mathrm{A1\&2}})$, estimates the intermediate precision of the daily means). For recovery assessment, data from sample A1&2 was considered. Table B5.1 presents the mean values and intermediate precision standard deviations of the analysis of these items. This table also reports the mean of the sediment mass of the spiked samples, $\bar{m}_{\mathrm{S}}$, in order to calculate the mean of spiked total As mass fraction, $\bar{w}_{\mathrm{S}}$ ($\bar{w}_{\mathrm{S}} = 3.761\ \mu\mathrm{g}/\bar{m}_{\mathrm{S}}$). The $\bar{w}_{\mathrm{S}}$ is 7.501 mg kg^{-1} and 7.591 mg kg^{-1} for spiked samples AS and CS, respectively.

Table B5.1. Means, $\bar{w}$, intermediate precision standard deviations, s_{I}, and number, *n*, of replicate determination of total As mass fractions of sediment samples A and C before (A1&2 and C) and after spiking (AS and CS) with 3.761 µg of As with a relative standard uncertainty of 0.92 %. Regarding the analysis of sample A, it is reported the performance of the mean of duplicates obtained under repeatability conditions. The mean mass of spiked samples, $\bar{m}_{\mathrm{S}}$, is also reported.

Sample	$\bar{w}$/ mg kg^{-1}	s_{I}/ mg kg^{-1}	$\bar{m}_{\mathrm{S}}$ (g)	*n*
A1&2	15.79	0.106	-	10
AS	22.546	0.782	0.5014	10
C	6.461	0.312	-	9
CS	14.31	1.83	0.4955	9
WQ-1	21.040	0.779	-	10

Note: $\bar{w}$, s_{I} and m_{S} are reported with the decimal places that allow a standard deviation with three significant figures. This number of figures is adequate for calculation precision.

The s_{I} difference between Samples C and CS suggests sediment heterogeneity is a major uncertainty component that should not be overlooked.

Equation B5.1 was used to calculate the mean analyte recovery and Eq.(B5.2) the respective standard uncertainty as a combination of Eq.(9) and (11).

$$\bar{\bar{R}} = \frac{\sum_{i=1}^{N}\bar{R}_i}{N} = \frac{\frac{\bar{w}_{\mathrm{RM}}}{W_{\mathrm{RM}}} + \frac{\bar{w}_{\mathrm{AS}} - \bar{w}_{\mathrm{A1\&2}}}{\bar{w}_{\mathrm{S(A)}}} + \frac{\bar{w}_{\mathrm{CS}} - \bar{w}_{\mathrm{C}}}{\bar{w}_{\mathrm{S(C)}}}}{3} = \\ = \frac{\frac{21.040}{23.00} + \frac{22.546 - 15.79}{7.501} + \frac{14.31 - 6.461}{7.591}}{3} = \\ = 0.9498 \qquad \text{(B5.1)}$$

where $\bar{\bar{R}}$ and $\bar{R}_i$ are the overall mean and the *i*-th mean analyte recovery, and $\bar{w}_{\mathrm{X}}$ and $\bar{w}_{\mathrm{Y(S)}}$ the estimated mean mass fractions of sample X and of sample Y spiking, respectively.

$$u_{\bar{\bar{R}}} = \sqrt{\left\{ \sqrt{\begin{gathered} \left(\frac{\bar{w}_{\mathrm{RM}}}{W_{\mathrm{RM}}}\right)^2 \left[\left(\frac{s_{\mathrm{I}}(w_{\mathrm{RM}})}{\bar{w}_{\mathrm{RM}}\sqrt{n}}\right)^2 + \left(\frac{U(W_{\mathrm{RM}})/k}{W_{\mathrm{RM}}}\right)^2\right] + \\ \left(\frac{\bar{w}_{\mathrm{AS}} - \bar{w}_{\mathrm{A1\&2}}}{\bar{w}_{\mathrm{S(A)}}}\right)^2 \left[\frac{\frac{s_{\mathrm{I}}^2(w_{\mathrm{AS}})}{n_{\mathrm{AS}}} + \frac{s_{\mathrm{I}}^2(w_{\mathrm{A1\&2}})}{n_{\mathrm{A1\&2}}}}{(\bar{w}_{\mathrm{AS}} - \bar{w}_{\mathrm{A1\&2}})^2} + u'^2_{\mathrm{S(A)}}\right] + \\ \left(\frac{\bar{w}_{\mathrm{CS}} - \bar{w}_{\mathrm{C}}}{\bar{w}_{\mathrm{S(C)}}}\right)^2 \left[\frac{\frac{s_{\mathrm{I}}^2(w_{\mathrm{CS}})}{n_{\mathrm{CS}}} + \frac{s_{\mathrm{I}}^2(w_{\mathrm{C}})}{n_{\mathrm{C}}}}{(\bar{w}_{\mathrm{CS}} - \bar{w}_{\mathrm{C}})^2} + u'^2_{\mathrm{S(C)}}\right] + \end{gathered}} \Bigg/ 3 \right\}^2 + s(\bar{R})^2} \tag{B5.2}$$

where $s_{\mathrm{I}}(w_{\mathrm{X}})$ and n_{X} are the intermediate precision standard deviation and number of determinations of sample X, $u'_{\mathrm{S(X)}}$ the relative standard uncertainty of sample spiking and $s(\bar{R})$ the standard deviation of the mean recoveries $\bar{R}$ estimated for the CRM and spiked samples A and C (i.e. 0.0732). The $u'_{\mathrm{S(X)}}$ is estimated from the combination of the relative standard uncertainty of As mass addition (0.92 %) and spiked sample mass (0.16 %) ($u'_{\mathrm{S(X)}} = (0.92\ \%^2 + 0.16\ \%^2)^{1/2} = 0.93\ \%$). The spiking can be easily planned to ensure $u'_{\mathrm{S(X)}}$ is negligible in order to simplify calculations. In Eq.(B5.2), the term $s(\bar{R})$ is added to avoid the need to assess if the $\bar{R}_i$ are metrologically different. Equation B5.3 presents the application of Eq.(B5.2) to the studied case.

$$u_{\bar{\bar{R}}} = \sqrt{\left\{ \sqrt{\begin{gathered} \left(\frac{21.040}{23.00}\right)^2 \left[\left(\frac{0.779}{21.040\sqrt{10}}\right)^2 + \left(\frac{1.84/2}{23.00}\right)^2\right] + \\ \left(\frac{22.546 - 15.79}{7.501}\right)^2 \left[\frac{\frac{0.782^2}{10} + \frac{0.106^2}{10}}{(22.546 - 15.79)^2} + 0.93\ \%^2\right] + \\ \left(\frac{14.31 - 6.461}{7.591}\right)^2 \left[\frac{\frac{1.83^2}{9} + \frac{0.312^2}{9}}{(14.31 - 6.461)^2} + 0.93\ \%^2\right] \end{gathered}} \Bigg/ 3 \right\}^2 + 0.0732^2} = 0.0800 \tag{B5.3}$$

The mean recovery is not significantly different from 1 given the $u_{\bar{\bar{R}}}$ as proven by the validity of the condition from Eq.(B5.4).

$$\begin{gathered} \frac{|1 - \bar{R}|}{u_{\bar{\bar{R}}}} \leq t_{\nu}^{0.95} \Leftrightarrow \\ \Leftrightarrow \frac{|1 - 0.9498|}{0.0800} \leq 4.30 \Leftrightarrow \\ \Leftrightarrow 0.6275 \leq 4.30 \end{gathered} \tag{B5.4}$$

Since the mean recovery uncertainty is dominated by a $s(\bar{R})$ with two degrees of freedom, it is assumed $u_{\bar{\bar{R}}}$ has the same number of degrees of freedom and $t_{\nu}^{0.95} = 4.30$ (value of the t-distribution for two degrees of freedom and 95% confidence level).

B5.7.iii. Additional uncertainty components
See subsection B4.7.iii.

B5.7.iv. Uncertainty combination and expansion:
Equations B5.5 to B5.6 present the algorithms used to calculate the expanded uncertainty, for approximately 95 % confidence level, for determinations of mass fraction w performed within Intervals I and II. Since $\bar{R}$ is equivalent to 1, it is assumed $\bar{R} = 1$ to calculate the relative standard uncertainty of the mean recovery, $u'_{\bar{R}}$ ($u'_{\bar{R}} = u_{\bar{R}}/1$) (see section 8.2.6.2).

Interval I [1.3 mg kg^{-1}; 6.46 mg kg^{-1}]:

$$\begin{aligned} U\langle \mathrm{I} \rangle &= 4.30\sqrt{s_{\mathrm{I}}^{2}\langle \mathrm{I} \rangle + \left(w \cdot \frac{u_{\bar{R}}}{1}\right)^{2}} = \\ &= 4.30\sqrt{0.292^{2} + (w \cdot 0.0800)^{2}} \end{aligned} \qquad \text{(B5.5)}$$

Interval II [6.46 mg kg^{-1}; 25 mg kg^{-1}]:

$$\begin{aligned} U\langle \mathrm{II} \rangle &= 4.30w\sqrt{s'^{2}_{\mathrm{I}}\langle \mathrm{II} \rangle + \left(\frac{u_{\bar{R}}}{1}\right)^{2}} = \\ &= 4.30w\sqrt{0.0452^{2} + 0.0800^{2}} = 0.395w \end{aligned} \qquad \text{(B5.6)}$$

Figure B5.1 presents the model of measurement, U, and target, U^{tg} (see section B4.4), expanded uncertainties for the measured mass fraction showing that only measurements below 9.5 mg kg^{-1} have fit-for-purpose measurement uncertainty. In this Example, the additional recovery data considered, when compared with Example B4, and the approximations performed produced an expanded uncertainty larger than that described in the previous example (Example B4).

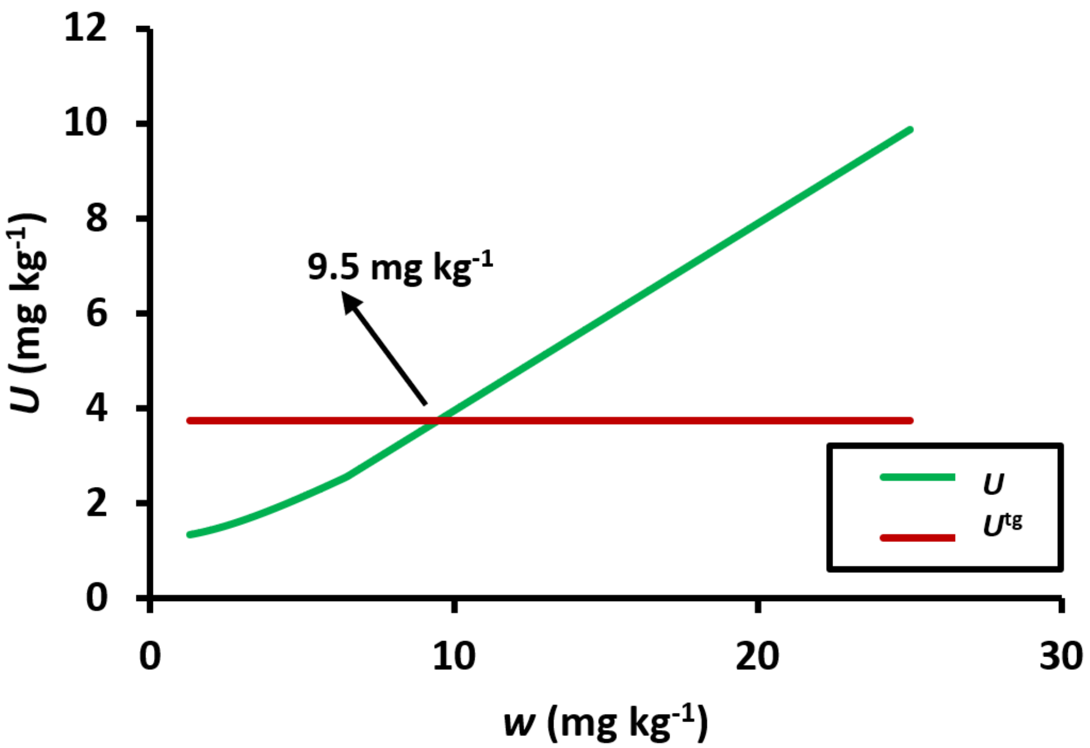

Figure B5.1. Model of measurement, U, and target, U^{tg}, expanded uncertainties, for 95 % confidence level, for the measured total As mass fraction, w, defined from the intermediate precision estimated from the analysis of three real marine sediments and analyte recovery determined from the analysis of a certified reference material and two spiked samples with native analyte. For mean analyte recovery uncertainty, it was considered the standard deviation of mean recoveries from the analysis of the three reference materials.

B5.7.v. Uncertainty contributions
See subsection B4.7.v.

B5.8. Uncertainty reduction

In the measurement uncertainty model described in subsection B5.7.iv., determinations above 9.5 mg kg^{-1} from Interval II have unacceptably large uncertainty. Since the standard uncertainty of mean analyte recovery is the major uncertainty component ($u'_{\bar{R}} = 0.0800$), it must be optimised to allow the determinations of measured values with adequately low uncertainty.

The mean analyte recovery uncertainty can be reduced by not taking the $s(\bar{R})$ term into account if mean recoveries estimated from the analysis of the CRM and spiked samples with native analyte are equivalent. A pair of mean recoveries, for instance those observed in the analysis of the CRM, $\bar{R}_{\mathrm{CRM}}$, and spiked samples A, $\bar{R}_{\mathrm{A}}$, are equivalent, for at the approximately 95 % confidence level, if the confidence interval of the mean recovery difference is equivalent to zero. This can be tested by the following condition:

$$\frac{|\bar{R}_{\mathrm{RM}} - \bar{R}_{\mathrm{A}}|}{\sqrt{u^2(\bar{R}_{\mathrm{RM}}) + u^2(\bar{R}_{\mathrm{A}})}} \leq 2 \Leftrightarrow \qquad \text{(B5.7)}$$

$$\Leftrightarrow \frac{|\bar{R}_{\mathrm{RM}}-\bar{R}_{\mathrm{A}}|}{\sqrt{\left(\frac{\bar{w}_{\mathrm{RM}}}{W_{\mathrm{RM}}}\right)^2\left[\left(\frac{s_{\mathrm{I}}(w_{\mathrm{RM}})}{\bar{w}_{\mathrm{RM}}\sqrt{n}}\right)^2+\left(\frac{U(W_{\mathrm{RM}})/k}{W_{\mathrm{RM}}}\right)^2\right]+\left(\frac{\bar{w}_{\mathrm{AS}}-\bar{w}_{\mathrm{A1\&2}}}{\bar{w}_{\mathrm{S(A)}}}\right)^2\left[\frac{\frac{s_{\mathrm{I}}^2(w_{\mathrm{AS}})}{n_{\mathrm{AS}}}+\frac{s_{\mathrm{I}}^2(w_{\mathrm{A1\&2}})}{n_{\mathrm{A1\&2}}}}{(\bar{w}_{\mathrm{AS}}-\bar{w}_{\mathrm{A1\&2}})^2}+u'^2_{\mathrm{S(A)}}\right]}}\le 2$$

Equation B5.8 details the application of Eq.(B5.7) for the comparison of the mean recovery estimated from the analysis of the CRM and spiked sampled A.

$$\frac{\left|\frac{21.040}{23.00}-\frac{22.546-15.79}{7.501}\right|}{\sqrt{\left(\frac{21.040}{23.00}\right)^2\left[\left(\frac{0.779}{21.040\sqrt{10}}\right)^2+\left(\frac{1.84/2}{23.00}\right)^2\right]+\left(\frac{22.546-15.79}{7.501}\right)^2\left[\frac{\frac{0.782^2}{10}+\frac{0.106^2}{10}}{(22.546-15.79)^2}+0.93\ \%^2\right]}}\le 2 \Leftrightarrow \Leftrightarrow 0.275\le 2 \tag{A5.8}$$

Since the condition is valid, the mean recoveries determined for the CRM and Sample A are equivalent. The same is observed when mean recoveries for CRM vs Sample C and Sample A vs Sample C are calculated with $|\bar{R}_i-\bar{R}_{\mathrm{C}}|/\sqrt{u^2(\bar{R}_{\mathrm{i}})+u^2(\bar{R}_{\mathrm{C}})}$ equal to 1.32 and 1.50, respectively (i = CRM or A). Therefore, Eq.(B5.9), equivalent to Eq. (B5.2) and (B5.3) without the $s(\bar{R})^2$ term, can be used to calculate the standard uncertainty of the mean analyte recovery since there is no need to include a component that expresses the differences in recovery for different samples. It should be remembered that performing three independent comparison for 95% confidence level of mean recoveries is only adequate if equivalence is concluded since there is a probability larger than 5 % of one of the three comparisons wrongly suggesting mean recoveries are different. If equivalence is not observed, a χ^2 test described in section 8.2.7 should be performed (See Example B6).

$$u_{\bar{\bar{R}}} = \sqrt{\left\{\left(\frac{21.040}{23.00}\right)^2\left[\left(\frac{0.779}{21.040\sqrt{10}}\right)^2+\left(\frac{1.84/2}{23.00}\right)^2\right]+\left(\frac{22.546-15.79}{7.501}\right)^2\left[\frac{\frac{0.782^2}{10}+\frac{0.106^2}{10}}{(22.546-15.79)^2}+0.93\ \%^2\right]+\left(\frac{14.31-6.461}{7.591}\right)^2\left[\frac{\frac{1.83^2}{9}+\frac{0.312^2}{9}}{(14.31-6.461)^2}+0.93\ \%^2\right]\right\}\Big/3} = 0.0323 \tag{B5.9}$$

For a $u_{\bar{\bar{R}}}=0.0323$, the $\bar{\bar{R}}$ is still equivalent to 1 ($\left|1-\bar{\bar{R}}\right|/u_{\bar{\bar{R}}}=1.554$) and the combined expanded uncertainty from a single analysis reduced as described by Eq.(B5.10) and (B5.11) for Intervals I and II:

Interval I [1.3 mg kg^{-1}; 6.46 mg kg^{-1}]:

$$U\langle \mathrm{I} \rangle = 2\sqrt{0.292^2 + (w \cdot 0.0323)^2} \tag{B5.10}$$

Interval II [6.46 mg kg^{-1}; 25 mg kg^{-1}]:

$$U\langle \mathrm{II} \rangle = 2w\sqrt{0.0452^2 + 0.0323^2} = 0.111w \tag{B5.11}$$

In this case, between 1.3 mg kg^{-1} and 25 mg kg^{-1} the *U* is smaller than the target value of 3.75 mg kg^{-1}. (see Figure B5.2).

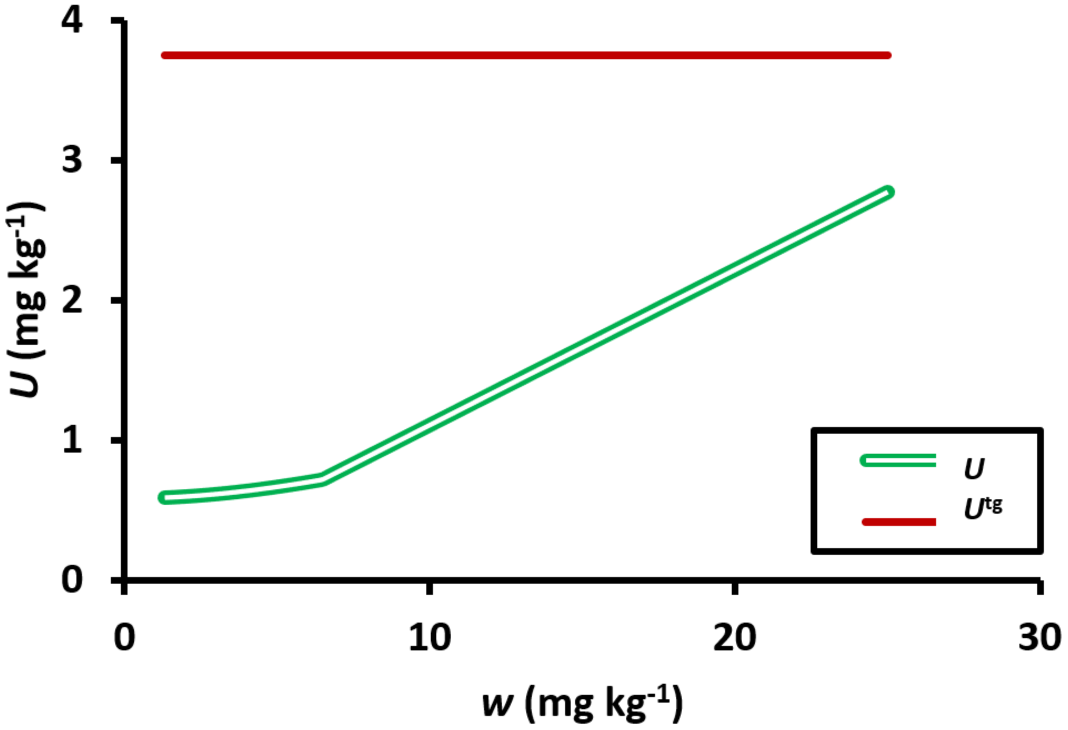

Figure B5.2. Model of measurement, *U*, and target, U^{tg}, expanded uncertainties, for 95 % confidence level, for the measured total As mass fraction, *w*, defined from the intermediate precision estimated from the analysis of three real marine sediments and analyte recovery determined from the analysis of a certified reference material and two spiked samples with native analyte. For the uncertainty associated with mean analyte recovery, the standard deviation of mean recoveries from the analysis of the three reference materials was not considered.

B5.9. Uncertainty model assumptions and limitations

The described measurement uncertainty model is applicable to the determination of total As, between 1.3 mg kg^{-1} and 25 mg kg^{-1}, for sediments not more heterogenous than analysed sediments A, B and C and affected by matrix effects equivalent to that observed from the analysis of the CRM WQB-1 and spikes samples A and C. The validity of the described measurement model can be checked through the analysis of other reference materials as described in section B1.9.

B5.10. References

See section **B4.9.**

B6: Determination of total As in sediments – Uncertainty evaluation III

Assessment of precision affected by sample heterogeneity from the analysis of homogenous and heterogeneous samples using Modelling 1 from Appendix A, and recovery estimated from the single analysis of various proficiency test samples

B6.1. Analytical method

See section B4.1.

B6.2. Stock solutions and sample conservation

See section B4.2.

B6.3. Measurement purpose

See section B4.3.

B6.4. Measurement requirements

See section B4.4.

B6.5. Analytical scope

See section B4.5.

B6.6. Specific feature of the example

This example describes a scenario where the heterogeneity of analysed samples can be a relevant uncertainty component. The measurement precision is estimated from the analysis of an homogeneous sample, sample heterogeneity is quantified from the duplicate analysis of several real samples and measurement trueness is assessed from the single analysis of several proficiency test samples.

B6.6.i. Precision uncertainty

The precision attributed to sample heterogeneity was quantified from results of the duplicate analysis of real samples, under repeatability conditions, and the intermediate precision was quantified from the analysis of a solution of 75 μg L^{-1} of As(III) (called DS), subject to the acid digestion and HGAAS quantification.

For an analytical portion of 0.5 g, 50 mL of 75 μg L^{-1} of As digested solution corresponds to 7.5 mg kg^{-1} of As in dry sediment. However, the mass of the analytical portion is not always exactly 0.5 g.

The results from the analysis of the DS do not express the precision of the measurement of the mass of the analytical portion (i.e., mass measurement performance) but this component is expressed in the duplicate analysis of real samples. The intermediate precision of the analysis of the DS also expresses the spiking precision (i.e., the precision of stock solution addition), but spiking was designed to make this component negligible.

B6.6.ii. Mean analyte recovery uncertainty

Analyte recovery was estimated from the single analysis of several proficiency test samples from different types of sediments.

Systematic effects from mass measurements are partially covered in the mean analyte recovery determinations and are negligible given the other uncertainty components.

B6.6.iii. Mean recovery correction

In this case, analyte recovery is estimated from the analysis of 22 sediments (See subsection B4.6.iii. for how to handle recovery deviations).

B6.7. Uncertainty calculations

B6.7.i. Precision uncertainty

Table B6.1 presents the results of the analysis of the DS performed on different days and Table B6.2 the results from the duplicate analysis of real sediment samples under repeatability conditions.

Table B6.1: HGAAS measurement results for a 75 µg L^{-1} As(III) solution after the acid digestion irradiated with microwaves. Determinations were performed, on different days, from independent digestions and spectrometer calibrations.

Estimated values, w/ mg L^{-1}					
76.05	75.10	88.27	87.20	73.73	83.27
77.20	80.85	70.57	92.12	75.06	76.55
79.13	73.06	78.16	86.99	73.24	77.83
74.08	73.00	77.92	66.17	76.22	85.24
80.07	74.74	76.55	79.66	85.32	80.49
$\bar{w}$/ mg L^{-1}	78.46	n	30		
$s_{\mathrm{I(DS)}}$/ mg L^{-1}	5.76	$s'_{\mathrm{I(DS)}}$	7.34 %		

$\bar{w}$ – Mean mass fraction; $s_{\mathrm{I(DS)}}$ and $s'_{\mathrm{I(DS)}}$ – mass fraction absolute and relative standard deviations.

The analytical interval between 0.5 mg kg^{-1} and 33 mg kg^{-1} (one fifth and 20 % above the lower and higher studied mass fractions from Samples S1359 and S1420, respectively)[*] is divided into two intervals: Interval I and II below and above 7.5 mg kg^{-1}. For the duplicate analysis of sediment samples reported in Table B6.2, when the mean value is below or above 7.5 mg kg^{-1}, the range of duplicate results is quantified as the absolute, $A_{\mathrm{(h)}}$, or relative, $A'_{\mathrm{(h)}}$, value. Seven and 23 ranges of duplicate results from Interval I and II were quantified. The mean of the absolute, $\bar{A}_{\mathrm{(h)}}$, and relative, $\bar{A}'_{\mathrm{(h)}}$, ranges are converted to absolute, $s_{\mathrm{r(h)}}$, and relative, $s'_{\mathrm{r(h)}}$, repeatability standard deviations for the analysis of heterogeneous samples after dividing by the factor 1.128 (See section 8.1.3)[†]. Therefore, the measurement precision models are:

Interval I [0.5 mg kg^{-1}; 7.5 mg kg^{-1}]:

$$\begin{gathered} s_{\mathrm{I}}\langle \mathrm{I}\rangle = \sqrt{\left(7.5 \cdot s'_{\mathrm{I(DS)}}\right)^2 + \left(\bar{A}_{\mathrm{(h)}}\langle \mathrm{I}\rangle/1.128\right)^2} \Leftrightarrow \\ \Leftrightarrow s_{\mathrm{I}}\langle \mathrm{I}\rangle = \sqrt{(7.5 \cdot 7.34\ \%)^2 + (0.2157/1.128)^2} \Leftrightarrow \\ \Leftrightarrow s_{\mathrm{I}}\langle \mathrm{I}\rangle = 0.5828\ \mathrm{mg\ kg^{-1}} \end{gathered} \tag{B6.1}$$

where $s'_{\mathrm{I(DS)}}$ is the relative standard deviation of the analysis of the DS (see Table B6.1) and $\bar{A}_{\mathrm{(h)}}\langle \mathrm{I}\rangle$ is reported in Table B6.2.

---

[*] For many applications of pragmatic uncertainty evaluations, absolute or relative uncertainty observed at a concentration level is applicable five times below or 20% above that level, respectively.

[†] If instead of the mean range of duplicate results, the difference of n duplicates, d_i, is considered, $s_{\mathrm{r}} = \sqrt{(\sum d_i^2)/2n}$.

Interval II [7.5 mg kg^{-1}; 33 mg kg^{-1}]:

$$\begin{gathered} s'_{\mathrm{I}}\langle \mathrm{II}\rangle = \sqrt{\left(s'_{\mathrm{I(DS)}}\right)^2 + \left(\bar{A}'_{(\mathrm{h})}\langle \mathrm{I}\rangle/1.128\right)^2} \Leftrightarrow \\ \Leftrightarrow s'_{\mathrm{I}}\langle \mathrm{II}\rangle = \sqrt{(7.341\ \%)^2 + (3.000\ \%/1.128)^2} \Leftrightarrow \\ \Leftrightarrow s'_{\mathrm{I}}\langle \mathrm{II}\rangle = 7.808\ \% \end{gathered} \tag{B6.2}$$

Table B6.2: Results from the replicate analysis of real sediment samples under repeatability conditions. For duplicate results w_1 and w_2 with a mean value below or above 7.5 mg kg^{-1} (Intervals I or II), absolute and relative ranges are estimated to calculated the absolute, $s_{\mathrm{r(h)}}$, and relative, $s'_{\mathrm{r(h)}}$, standard deviations, respectively.

	Measured values			Absolute range, $A_{(\mathrm{h})}$/ mg kg^{-1}	Relative range, $A'_{(\mathrm{h})}$/ (%)
Sample	w$_1$/ mg kg^{-1}	w$_2$/ mg kg^{-1}	Interval		
S0560	5.93	5.74	Interval I	0.19	-
S0498	4.36	4.01	Interval I	0.35	-
S0789	13.66	13.65	Interval II	-	0.07
S0239	15.77	16.46	Interval II	-	4.28
S0240	8.51	8.15	Interval II	-	4.32
S0379	10.19	10.40	Interval II	-	2.04
S0240	8.19	8.16	Interval II	-	0.37
S0821	6.25	6.34	Interval I	0.09	-
S0831	8.16	7.68	Interval II	-	6.06
S0841	13.16	13.00	Interval II	-	1.22
S0379	10.07	9.65	Interval II	-	4.26
S0239	14.21	14.01	Interval II	-	1.42
S0239	17.94	18.67	Interval II	-	3.99
S0845	9.67	9.23	Interval II	-	4.66
S0897	21.15	20.72	Interval II	-	2.05
S0906	19.64	20.02	Interval II	-	1.92
S0924	2.81	2.64	Interval I	0.17	-
S0932	11.11	11.03	Interval II	-	0.72
S0952	10.00	10.11	Interval II	-	1.09
S1359	2.30	2.25	Interval I	0.05	-
S0897	17.75	16.81	Interval II	-	5.44
S0913	8.59	8.01	Interval II	-	6.99
S1368	20.10	21.67	Interval II	-	7.52
S1369	11.98	11.98	Interval II	-	0.00
S1420	27.85	27.23	Interval II	-	2.25
S0931	5.58	5.85	Interval I	0.27	-
S0958	7.37	7.71	Interval II	-	4.51
S0951	4.62	4.23	Interval I	0.39	-
S1368	20.13	19.78	Interval II	-	1.75
S1369	11.79	11.54	Interval II	-	2.14
			Mean	$\bar{A}_{(\mathrm{h})} =$ $= 0.2157$ mg kg^{-1}	$\bar{A}'_{(\mathrm{h})} =$ $= 3.00$ %
			n	7	23
			$s_{\mathrm{r(h)}}$	0.1912 mg kg^{-1}	-
			$s'_{\mathrm{r(h)}}$	-	2.66 %

B6.7.ii. Mean analyte recovery uncertainty

The mean analyte recovery was evaluated from the results from the single analysis of 22 proficiency test samples. Table B6.3 presents the results from the participation in proficiency tests, PT, where column R and $u(R)$ report the mean analyte recovery (ratio between the measured, *w*, and reference value, *W*) and their standard uncertainty. Equations B6.3 and B6.4 present the calculation of R and u_R for analysis of the first sediment sample from PT2014-1 based on Eq.(8) and (9):

$$R = \frac{w}{W} = \frac{12.000}{14.460} = 82.99\ \% \qquad \text{(B6.3)}$$

$$\begin{aligned} u_R &= \sqrt{\left(\frac{w}{W}\right)^2 \left[\left(\frac{s_\mathrm{I}(w)}{w\sqrt{n}}\right)^2 + \left(\frac{u(W)}{W}\right)^2\right]} \Leftrightarrow \\ \Leftrightarrow u_R &= \sqrt{(0.8299)^2 \left[\left(\frac{0.8808}{12.000\sqrt{1}}\right)^2 + \left(\frac{0.2340}{14.460}\right)^2\right]} = 0.06238 \end{aligned} \qquad \text{(B6.4)}$$

where $s_\mathrm{I}(w)$ is intermediate precision standard deviation from the analysis of the PT sample estimated from the developed precision model where the sample heterogeneity component is not considered since PT samples are homogeneous. Since *w* is larger than 7.5 mg kg^{-1}, the precision model for Interval II applicable to homogeneous samples should be considered: $s'_{\mathrm{I(homo)}}\langle \mathrm{II}\rangle = 7.34\ \%$. Therefore, for the first sample of PT2014-1, $s_{\mathrm{I(homo)}}\langle \mathrm{II}\rangle = 12.000\ \mathrm{mg\ kg^{-1}} \cdot 7.34\ \% = 0.8808\ \mathrm{mg\ kg^{-1}}$.

The mean analyte recovery, $\bar{\bar{R}}$, estimated from the analysis of the 22 PT samples is calculated from Eq.(B6.5) or the mean of the values of column *R* from Table B6.3.

$$\bar{\bar{R}} = \frac{\sum_{i=1}^{N} \frac{w_i}{W_i}}{N} = 0.9174 \qquad \text{(B6.5)}$$

where *N* is 22 and *i* = 1 to 22, the PT index. Equation B6.6 describes the quantification of standard uncertainty of mean recovery, $u_{\bar{\bar{R}}}$.

$$\begin{aligned} u_{\bar{\bar{R}}} &= \sqrt{\frac{\sum_{i=1}^{N} \left(\frac{w_i}{W_i}\right)^2 \left[\left(\frac{s_\mathrm{I}(w_i)}{w_i\sqrt{n_i}}\right)^2 + \left(\frac{u(W_i)}{W_i}\right)^2\right]}{N^2} + s(R)^2} \Leftrightarrow \\ \Leftrightarrow u_{\bar{\bar{R}}} &= \sqrt{0.01585^2 + 0.04607^2} = 0.04872 \end{aligned} \qquad \text{(B6.6)}$$

where $s(R)$ is the standard deviation of all *R* values, considered to avoid the need to check the equivalence of recoveries observed in the various PT.

In order to assess if recoveries are equivalent, χ^2 is determined using Eq.(16.2):

$$\chi^2 = \sum_{k=1}^{N}\left(\frac{R_k-\bar{\bar{R}}}{u(R_k)}\right)^2 = \begin{array}{c}\left(\dfrac{0.8299-0.9174}{0.0624}\right)^2 \\ +\displaystyle\sum_{k=2}^{21}\left(\frac{R_k-\bar{\bar{R}}}{u(R_k)}\right)^2+ \\ \left(\dfrac{1.0218-0.9174}{0.0817}\right)^2 \\ =8.80\end{array} \qquad \text{(B6.7)}$$

Since, $\chi^2 = 8.80$ is smaller than $\chi^2_{0.05,22-1} = 32.7$, it can be concluded that recovery values are equivalent and the term $s(R)$ can be removed from Eq.(B6.6). If this term is removed, $u_{\bar{\bar{R}}} = 0.01585$ and $|1-\bar{\bar{R}}|/u_{\bar{\bar{R}}} = |1-0.9174|/0.01585 = 5.21$, requiring the correction of original measured values, w, for observed low recovery ($w_\mathrm{C} = w/\bar{\bar{R}}$; where w_C is the mass fraction corrected for observed recovery). Alternatively, it can be decided to keep an overestimated $u_{\bar{\bar{R}}} = 0.04872$ that gives a $|1-\bar{\bar{R}}|/u_{\bar{\bar{R}}} = 1.70$ that does not require recovery corrections.

Table B6.3: Results from the single determination of total As of sediment samples from various proficiency tests.

PT	w/ mg kg^{-1}	W/ mg kg^{-1}	$u(W)$ / mg kg^{-1}	R	u_R
R2014-1	12.000	14.460	0.2340	0.8299	0.0624
R2014-1	7.000	7.396	0.1904	0.9465	0.0783
R2014-2	6.900	7.495	0.2254	0.9206	0.0785
R2014-2	6.600	7.213	0.1951	0.9150	0.0802
R2015-1	12.800	14.220	0.4887	0.9001	0.0730
R2015-1	19.200	21.360	0.7165	0.8989	0.0726
R2015-2	17.400	18.192	0.4227	0.9565	0.0736
R2015-2	8.130	8.299	0.1992	0.9796	0.0757
R2016-1	13.400	14.875	0.5777	0.9009	0.0748
R2016-1	16.200	18.442	0.7851	0.8784	0.0745
R2016-2	7.654	8.360	0.2457	0.9156	0.0724
R2017-1	6.636	7.160	0.2136	0.9268	0.0817
R2017-1	8.496	9.440	0.3262	0.9000	0.0730
R2017-2	18.600	22.100	0.2363	0.8416	0.0624
R2017-2	10.000	11.500	0.3129	0.8696	0.0681
R2018-1	12.800	14.100	0.5563	0.9078	0.0757
R2018-1	10.100	11.300	0.4900	0.8938	0.0762
R2018-2	17.400	17.800	0.3002	0.9775	0.0736
R2018-2	8.240	8.720	0.3258	0.945	0.0778
R2019-1	16.500	17.000	0.4053	0.9706	0.0749
R2019-1	8.480	9.580	0.2885	0.8852	0.0702
R2019-2	11.700	11.450	0.3637	1.0218	0.0817

w – measured value; W – reference value; $u(W)$ – standard uncertainty of W;
R – analyte recovery; u_R – Standard uncertainty of R.

B6.7.iii. Additional uncertainty components
See subsection B4.7.iii.

B6.7.iv. Uncertainty combination and expansion

Equations B6.8 to B6.9 present the calculation of the expanded uncertainty from single analysis performed in Interval I and II, set for measured values before recovery correction, w.

Interval I [0.5 mg kg^{-1}; 7.5 mg kg^{-1}]:

$$\begin{aligned} U\langle \mathrm{I} \rangle &= 2\sqrt{s_{\mathrm{I}}^2\langle \mathrm{I} \rangle + \left(\frac{w \cdot u_{\bar{\bar{R}}}}{\bar{\bar{R}}^2}\right)^2} = \\ &= 2\sqrt{0.5828^2 + [(w \cdot 0.01585)/0.9174^2]^2} \end{aligned} \qquad \text{(B6.8)}$$

Interval II [7.5 mg kg^{-1}; 25 mg kg^{-1}]:

$$\begin{aligned} U\langle \mathrm{II} \rangle &= 2\frac{w}{\bar{\bar{R}}}\sqrt{s'^2_{\mathrm{I}}\langle \mathrm{II} \rangle + \left(\frac{u_{\bar{\bar{R}}}}{\bar{\bar{R}}}\right)^2} = \\ &= 2\frac{w}{\bar{\bar{R}}}\sqrt{0.07808^2 + (0.01585/0.9174)^2} = 0.160\frac{w}{\bar{\bar{R}}} \end{aligned} \qquad \text{(B6.9)}$$

B6.7.v. Uncertainty contributions

See subsection B4.7.v.

B6.8. Uncertainty reduction

According to the measurement model described in section B6.7iv, when w is larger than 21.5 mg kg^{-1} (w_{C} > 23.4 mg kg^{-1}), in Interval II, $U\langle \mathrm{II} \rangle$ is larger than the U^{tg} of 3.75 mg kg^{-1}. Therefore, a third interval must be defined, Interval III, for w between 21.5 mg kg^{-1} and 29.7 mg kg^{-1} (w_{C} = 32.4 mg kg^{-1}) reducing interval II to [7.5 mg kg^{-1}; 21.5 mg kg^{-1}]. In interval III ([21.5 mg kg^{-1}; 29.7 mg kg^{-1}], excluding 21.5 mg kg^{-1}), the samples must be analysed twice on two different days. The measurement uncertainty is calculated using the algorithm of Eq.(B6.10).

$$\begin{aligned} U\langle \mathrm{III} \rangle &= 2\frac{w}{\bar{\bar{R}}}\sqrt{\frac{s'^2_{\mathrm{I}}\langle \mathrm{II} \rangle}{2} + \left(\frac{u_{\bar{\bar{R}}}}{\bar{\bar{R}}}\right)^2} = \\ &= 2\frac{w}{\bar{\bar{R}}}\sqrt{\frac{0.07808^2}{2} + (0.01585/0.9174)^2} = 0.1157\frac{w}{\bar{\bar{R}}} \end{aligned} \qquad \text{(B6.10)}$$

The upper limit of interval III for w was reduced from 33 mg kg^{-1} to 29.7 mg kg^{-1} because for duplicate analysis performed on different days the expanded uncertainty gets larger than the target value of 3.75 mg kg^{-1} above 29.7 mg kg^{-1}.

Figure B6.1 presents the measurement uncertainty model based on three intervals that allows measurements fit for assessing the conformity with the Portuguese law for contamination of sediments.

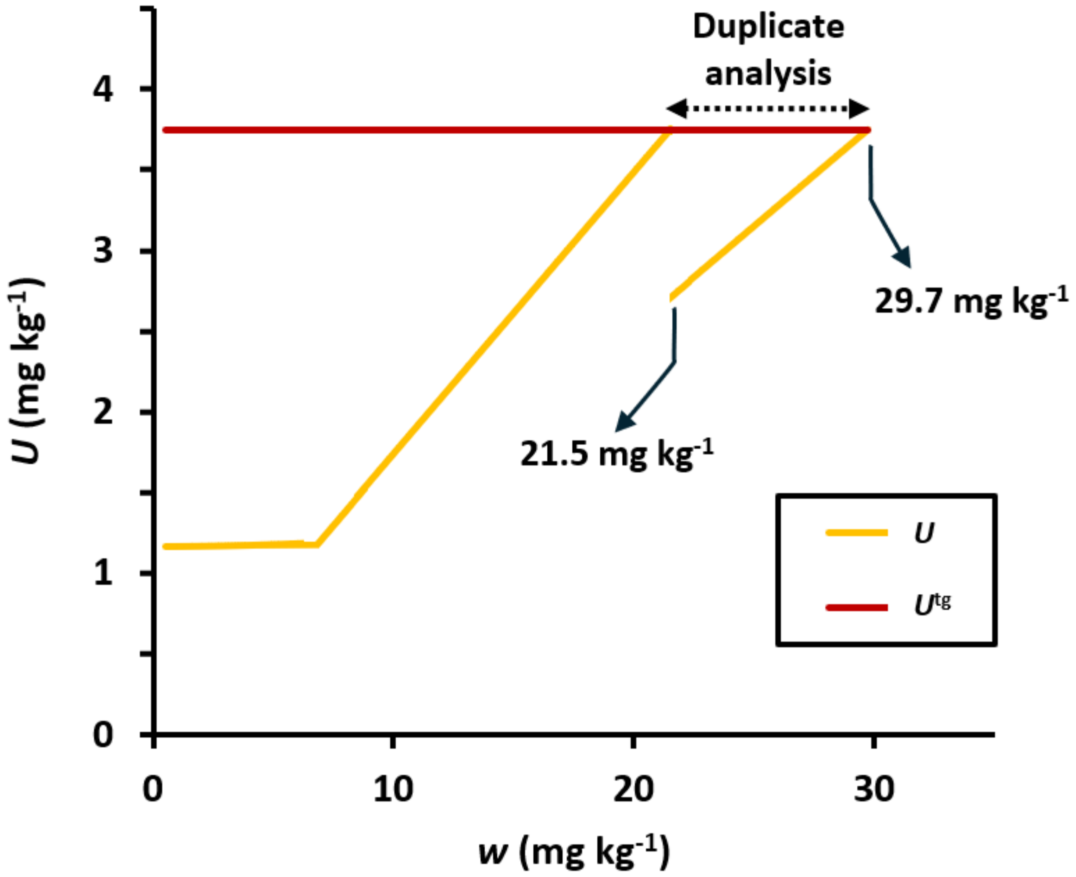

Figure B6.1. Model of measurement, *U*, and target, U^{tg}, expanded uncertainties, for 95 % confidence level, for the measured total As mass fraction, *w*, defined from the precision estimated from the analysis of homogenous and heterogenous samples and analyte recovery determined from the single analysis of samples from 22 proficiency tests. Above 23.4 mg kg^{-1}, the sample is analysed in duplicate under intermediate precision conditions.

B6.9. Uncertainty model assumptions and limitations

The described measurement uncertainty model is applicable to the determination of total As, between 0.54 mg kg^{-1} and 32.4 mg kg^{-1} (results corrected for recovery), for sediments not more heterogenous than the sediments analysed in the study and affected by matrix effects equivalent to that observed from the analysis of 22 sediments from proficiency tests.

B6.10. References

See section **B4.9.**

[B6.1] Carla Palma, “Determination of metals in sediments”, Lisboa: Instituto Hidrográfico, 2015.

Conclusions from examples B4, B5 and B6

Figure B7.1 presents the optimised models of the expanded uncertainty, at the 95 % confidence level, for the determination of total As in sediments described in Examples B4, B5 and B6. These models allow determinations with an expanded uncertainty lower than the target (maximum admissible) value of 3.75 mg kg^{-1}. Example B4 differs from Example B5 in the data used for evaluating analyte recovery. In Example B4 data from the analysis of one Certified Reference Material is used, whereas Example B5 includes additional data from two spiked samples with native analyte. The option to assess analyte recovery from more reference materials in Example B5 did not affect the evaluated measurement uncertainty significantly. In Example B6, measurement precision and analyte recovery were evaluated from data different from that used in Examples B4 and B5. This had an impact on the analytical range and reported uncertainty. The measurement model from Example B6 is based on a larger diversity of sediment samples for assessing both measurement precision and analyte recovery. This suggests it is more likely to be adequate for a wider range of sediment types. Example B6 also differs from the other examples due to the fact that correction of measured values for observed recovery is required.

All three described uncertainty evaluation are technically correct but can have different applicability regarding sediments heterogeneity and analysis matrix effects.

After initial development, the adequacy of measurement uncertainty models should be regularly monitored to assess the need to improve uncertainty components estimated in relation to the type of sediments characterised.

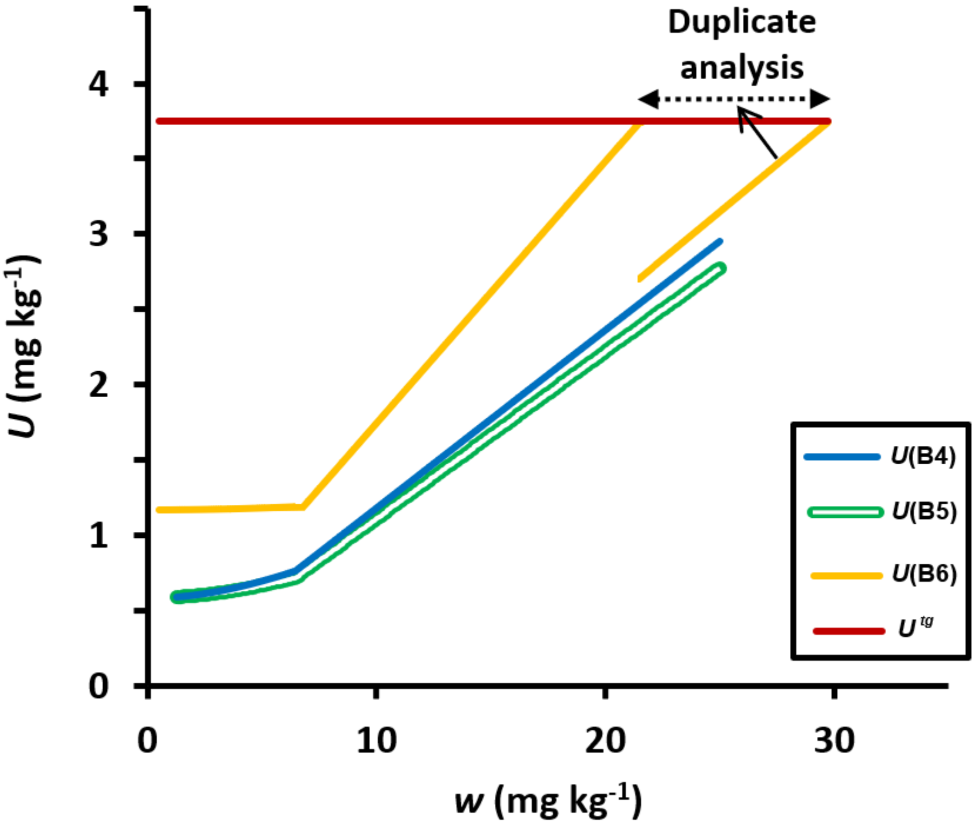

Figure B7.1. Models of measurement, U, and target, U^{tg}, expanded uncertainties, for 95 % confidence level, for the measured total As mass fraction of sediments described in Examples B4, B5 and B6.

Appendix C. Allowance for acceptable recovery intervals

C1.1. Application of acceptable recovery intervals

The checks and calculations in sections 8.2.3 to 8.2.6.2 provide a simplified approach to uncertainty evaluation when recovery or apparent recovery are checked for significant differences from complete recovery of the analyte of interest; that is, the calculations assume that the test method is checked for recovery significantly different from 100 %. Often, however, regulations or validation procedures permit an analytical procedure to be used subject to measured recovery being within an acceptable interval without requiring a correction for limited analyte recovery as long as the measured recovery is within the permitted interval. This is particularly common where perfect recovery is either impractical or unnecessary for particular purposes, and where laboratories have free choice of measurement method subject to validation constraints on precision and recovery.

This section gives brief information on some of the different criteria which can be used to confirm that recovery is within acceptable limits, together with guidance on the implications for reported measurement uncertainty.

C1.2. General criteria for acceptable recovery

Given acceptable limits for analyte recovery, three general criteria can be used:

i) *The recovery is within acceptable limits, the risk of wrong acceptance of unsuitable method is low*

   This corresponds to 'stringent acceptance' in the Eurachem guide on conformity with specified requirements [C1]. A common test for this purpose is the Two One-Sided t Test (TOST) [C2] in which the measured recovery is checked against the upper and lower limits of the acceptable interval, using the one-tailed value of Student's *t*, and the method is accepted if the measured recovery is demonstrably inside the acceptable interval with at least 95 % confidence. A slightly more stringent test uses the expanded uncertainty with coverage factor $k = 2$, again accepting only if the uncertainty interval for the measured recovery is entirely inside the acceptable interval.

ii) *The measured recovery is within the acceptable interval*

   This corresponds to 'simple acceptance' in the Eurachem guide on conformity with specified requirements [C1]. The method is accepted if the measured recovery value is within the acceptable interval, without regard to any uncertainty in the recovery estimate.

iii) *The recovery is within acceptable limits, the risk of wrong rejection of suitable method is low*

   This corresponds to 'relaxed acceptance' in the Eurachem guide on conformity with specified requirements [C1]. A typical test for this purpose compares the measured recovery with each limit, using the one-tailed value of Student's *t*, and the method is rejected only if the measured recovery is demonstrably outside the acceptable interval with at least 95 % confidence. A slightly less stringent test uses the expanded uncertainty with $k = 2$, rejecting only if the uncertainty interval for the measured recovery is entirely outside the acceptable interval.

These different criteria are illustrated in Figure C1. An important feature of the different criteria is that they lead to appreciably different intervals within which measured recovery would be considered acceptable.

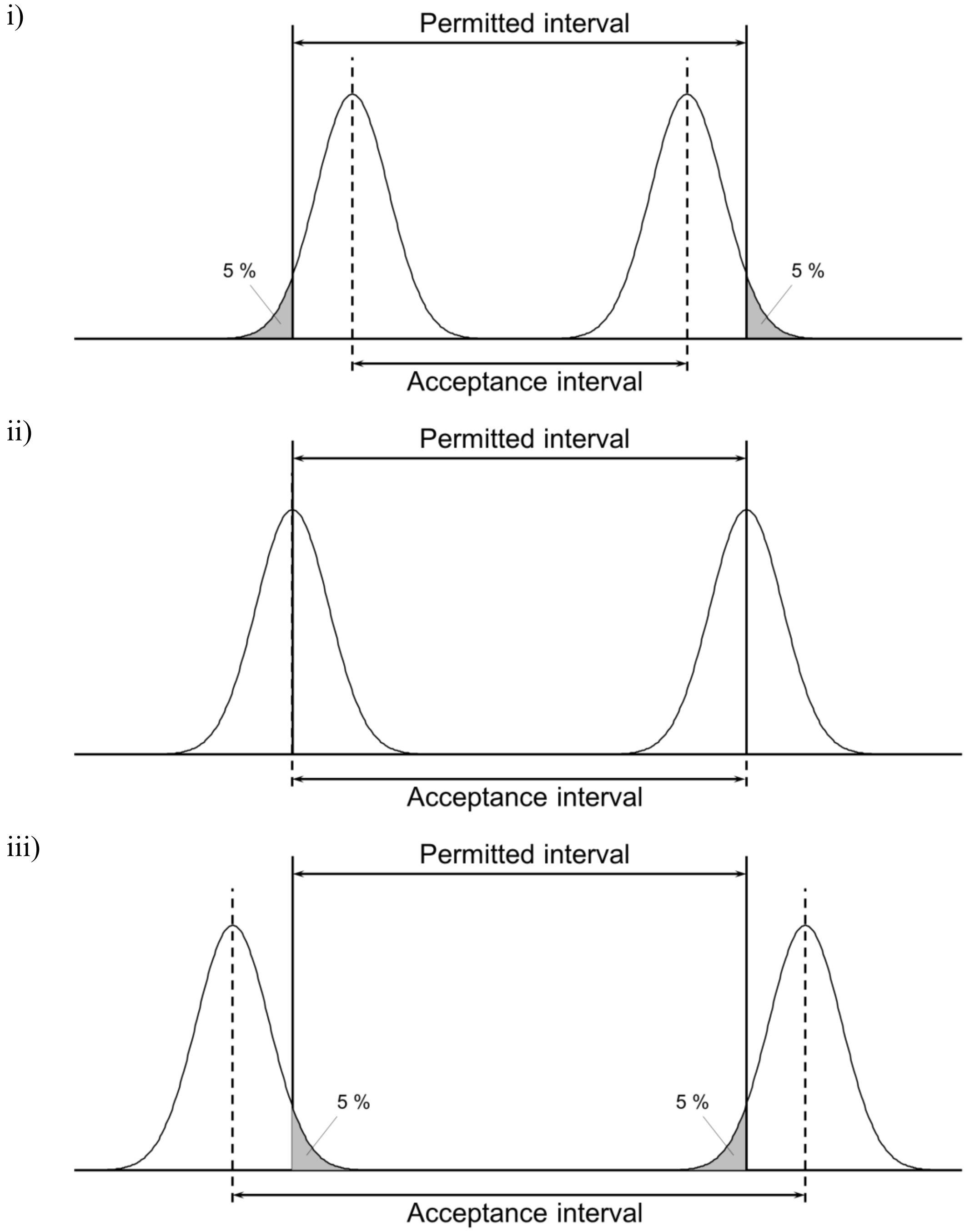

Figure C1. The figure illustrates the different criteria for observed recovery. i) Stringent acceptance, using the TOST procedure; ii) Simple acceptance; iii) Relaxed acceptance, using one-sided tests. Solid vertical lines show the lower and upper limits of the permitted interval of recoveries; dashed vertical lines show possible measured values that just meet the acceptability criteria. The permitted interval is the interval of recoveries permitted by regulation or laboratory policy. The distribution curves show the distribution of possible true recovery values inferred from the measured value(s). The acceptance interval is the interval within which a measured (mean) recovery would be considered acceptable.

C1.3. Uncertainty allowance for acceptable recovery interval

As Figure C1 shows, as the wideness of acceptance intervals increases, the stringency of the recovery control decreases. In practice, each different criterion will result in acceptance of a different interval of possible recoveries, which would each lead to some, uncorrected, bias in measurement on test samples.

In these circumstances, a conservative and simple approach is to include an additional component of uncertainty to allow for the possible bias arising from the permitted interval of recovery. A common and conservative allowance is to take the interval of recoveries that might reasonably be accepted as the limits of a rectangular distribution, and to include the standard uncertainty derived from this as an additional contribution to the recorded measurement uncertainty. For the criteria above, and assuming the recovery limits are symmetric about 100 % recovery, the limits of the required rectangular distribution are just the limits of the acceptance interval in each case. This follows primarily because the test criteria use symmetric distributions that vary little with location.

For permitted recovery limits that are symmetric about 100 % recovery, therefore, a conservative additional allowance can be calculated as follows:

i) Determine the width Δ_a of the acceptance interval corresponding to the test criterion in use.
ii) Calculate the standard uncertainty u_a corresponding to a rectangular distribution with width Δ_a as

$$u_\mathrm{a} = \frac{\Delta_a}{2\sqrt{3}} \tag{C1}$$

The width Δ_a is calculated from the width of the permitted interval, adjusted for the particular coverage factor or critical value used to determine whether a measured recovery complies with the permitted interval of recovery. Let the coverage factor or critical value be k_a, which may be, for example, 1.64 in the case of one-sided tests or 2.0 if using an expanded uncertainty with coverage factor $k = 2$. Then:

$$\Delta_\mathrm{a} = \begin{cases} \Delta_\mathrm{P} - 2k_a u_R, & \text{For stringent acceptance (case i)} \\ \Delta_\mathrm{P}, & \text{For simple acceptance (case ii)} \\ \Delta_\mathrm{P} + 2k_a u_R, & \text{For relaxed acceptance (case iii)} \end{cases} \tag{C2}$$

where Δ_P is the width of the permitted interval and u_R the standard uncertainty associated with the measured recovery.

Example C1: Uncertainty from the allowance for acceptable recovery interval

A regulation for measurement of contaminant metals in soil permits recovery to be (100 ± 10) %, using relaxed acceptance (criterion iii above) with a one-tailed hypothesis test. The laboratory measures apparent recovery as the mean of six replicate observations on a certified reference material, giving a relative uncertainty $u_{\bar{\mathrm{R}}}$ (from Eq. (9)) of 2.4 %, with (for clarity) large degrees of freedom. The interval (100 ± 10) % gives the width of the permitted interval as $\Delta_\mathrm{P} = 110 - 90 = 20\ \%$. For a one-side test at 95 % confidence, $k_\mathrm{a} = 1.64$, then, applying Eq. (C2) where u_R is substituted by $u_{\bar{\mathrm{R}}}$ for a mean recovery:

$$\Delta_\mathrm{a} = \Delta_\mathrm{P} + 2k_\mathrm{a}u_{\bar{R}} = (20 + 2 \times 1.64 \times 2.4) = 27.87\ \%$$

From this, the additional allowance u_a for permitted variations in recovery is

$$u_\mathrm{a} = \frac{\Delta_\mathrm{a}}{2\sqrt{3}} = \frac{27.87}{2 \cdot 1.73} = 8.1\ \%$$

This can be combined with the uncertainty $u_{\bar{R}}$ associated with the determination of recovery, and subsequently with precision contributions, to provide a conservative statement of uncertainty that allows for permitted variation in recovery.

C1.4. Asymmetric permitted interval for recovery

Permitted recovery intervals frequently recognise the fact that true analyte recovery should not exceed 100 %, and are set to permit a greater net loss of analyte. For example, mean measured recovery for pesticide analysis is often permitted to be in the interval 70 % to 120 %, without applying a correction. Such an interval implicitly accepts a modest loss of analyte; here, the permitted interval is centred on 95 % recovery.

There remains no broadly accepted approach to adjusting statements of uncertainty for procedures that result in a consistent, uncorrected bias. However, it is clear that use of equations (C1) and (C2) above, risk an understatement of the uncertainty, because the most extreme permitted recovery is further from 100 % than the half-width of the permitted interval.

In such a case, where no correction for recovery is expected, options include:

- Reporting observed recovery to the client;
- Further increasing the uncertainty reported with the results.

Possible approaches to handle these situations are:

i) A simple, but conservative, adjustment to the uncertainty contribution of Eq. (C1) is to replace $\Delta_a/2$ with the maximum permitted deviation from 100 %. In the pesticide example immediately above, and assuming simple acceptance (criterion ii), this would replace $(120 - 70)/2$ with $(100 - 70)$.

ii) The variance of an asymmetric rectangular distribution about a given central recovery value, R_C (ex. 95% in the pesticide example) can be calculated from Eq.(C1) combined with the squared offset δ_asym; that is, by replacing u_a from Eq.(C1) with $\sqrt{u_a^2 + \delta_\mathrm{asym}^2}$ where δ_asym is the difference between the mean observed recovery and R_C. This leads to a smaller allowance than the very crude approximation implied by taking the width as twice the maximum deviation.

Other approaches to handle such cases have been proposed [C3, C4].

C1.5. References

[C1] A. Williams and B. Magnusson (eds.), Eurachem/CITAC Guide: Use of uncertainty information in compliance assessment (2nd ed.), Eurachem, 2021. ISBN 978-0-948926-38-9.

[C2] D. J. Schuirmann, “A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability”, *J. Pharmacokinet. Biopharm.*, vol. 15, pp. 657–680, 1987.

[C3] B. Magnusson and S. L. R. Ellison, “Treatment of uncorrected measurement bias in uncertainty estimation for chemical measurements”, *Anal. Bioanal. Chem.*, vol. 390, pp. 201–213, 2008.

[C4] G. E. O’Donnell and D. B. Hibbert, “Treatment of bias in estimating measurement uncertainty”, *Analyst*, vol. 130, pp. 721–729, 2005.

ISBN 978-972-9348-30-3