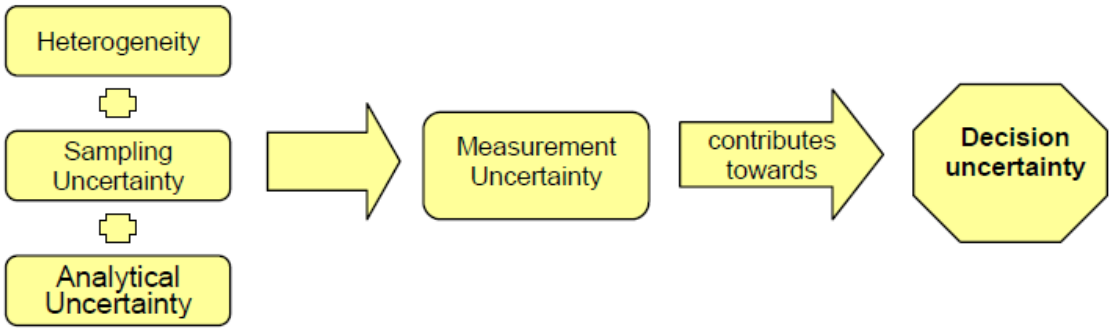
UNCERTAINTY FROM SAMPLING

- A NORDTEST HANDBOOK FOR SAMPLING PLANNERS ON SAMPLING QUALITY ASSURANCE AND UNCERTAINTY ESTIMATION

Based on the *Eurachem Guide Measurement uncertainty arising from sampling - A guide to methods and approaches*



By

Bertil Magnusson, Mikael Krysell,

Eskil Sahlin and Teemu Näykki

Nordtest
Slettetoften 9,
DK-2630 Taastrup

Telephone +45 21 21 44 44
nordtest@nordtest.info
www.nordtest.info

Uncertainty from sampling

A Nordtest Handbook for sampling planners on sampling quality assurance and uncertainty estimation

Based on the Eurachem Guide *Measurement uncertainty arising from sampling – A guide to methods and approaches*

2nd edition, June 2020

Bertil Magnusson, Trollboken AB, SE
Mikael Krysell, Swedish Agency for Marine and Water Management, SE
Eskil Sahlin, RISE Research Insititutes of Sweden, SE
Teemu Näykki, Finnish Environment Institute, FI

Authors of the 1st edition 2007
Christian Grøn, Jette Bjerre Hansen, Bertil Magnusson, Mikael Krysell, Ulla O. Lund, Kirsten Jebjerg Andersen and Astrid Nordbotten

Valuable comments on the 2nd edition were provided by

David Milde, CZ
Elvar Theodorsson, SE
Mike Ramsey, UK
Peter Rostron, UK
Stephen Ellison, UK
Wolfhard Wegsheider, AU
Ulf Örnemark, SE
and many devoted colleagues

Downloading address

www.nordtest.info

Recommended citation

This document should be cited* as "B. Magnusson, M. Krysell, E. Sahlin and T. Näykki, Uncertainty from sampling, Nordtest Report TR 604 (2nd) 2020, ISBN 978-91-89167-31-5. Available from www.nordtest.info."
*Subject to journal requirements

English edition

Second edition 2020

ISBN 978-91-89167-31-5

Foreword

This handbook provides practical guidance on sampling uncertainty estimation in the Nordtest TR handbook format. The handbook is an extract of, and based on the principles, methods and text of the *Eurachem Guide Estimation of measurement uncertainty arising from sampling*. The Eurachem guide is more extensive and provides details on theory and additional examples. In 2019 Eurachem published a revised edition and therefore also the Nordtest handbook is revised.

The major changes in this second edition are:

- the possible use of an *unbalanced design* is mentioned in Section 6 to estimate sampling uncertainty more cost-effectively;
- a detailed discussion in Section 7 on uncertainty contributions included and neglected using the experimental designs in this handbook;
- introduction in Section 9.6 using log-transformed data and an uncertainty factor, ${}^{F}U$, to handle expanded uncertainties of more than 30 %;
- a recommended software, RANOVA (in Excel), to be used for ANOVA calculations;
- updates to definitions and references to reflect current.

The overall purpose of this handbook is to provide a set of tools for calculation and control of the sampling uncertainty of the sampling procedure. It is the intention to make these tools and the understanding of their use available outside the world of analytical chemistry, although the basic principles applied originate from analytical chemistry. We hope that this is achieved but if not, please recall that statistics generally seem to be more complicated than it actually is…

How to use the Handbook

This Handbook was prepared as a helping hand for those who want to understand and control the uncertainty of the sampling part of their investigation, monitoring or control programmes.

The background, theory and principles are described in the text. and worked examples are given as annexes. The emphasis is simple explanations, with text boxes giving specific guidance and justification for the procedures. Figures illustrate the points made, and example boxes show the principles and the calculations:

Section 1 is a description of the scope of the Handbook;

Section 2 gives a list of abbreviation and symbols used;

Sections 3 and 4 provide the context of sampling as part of a measurement process with a purpose and a requirement for a defined quality;

Section 5 describes sources of error and uncertainty in sampling;

Section 6 describes the experimental designs used for estimating sampling uncertainty as well as measurement uncertainty;

Section 7 discusses which contributions to measurement uncertainty that are included and neglected with the experimental designs;

Section 8 describes how sampling validation and quality control can be designed;

Section 9 is the main section and here we present the statistical methods that enable calculation of measurement, sampling and analytical uncertainty.

The examples in the annexes illustrate the application of different statistical methods and tools, while allowing you to follow all steps of the calculations. Although the examples are given for specific matrices (groundwater, iron ore, baby food and wastewater) the approaches are widely applicable.

Annex	Matrix	Application	Sampling validation	QC	Design	Calculations
A	Groundwater	Monitoring	X	X	Balanced double split	Relative range
B	Iron ore	Production	X	-	Balanced triple split	Absolute range
C	Baby food	Surveillance	X	X	Balanced double split	ANOVA
D	Wastewater	Surveillance	-	-	Time series	Variography

Annex E contains the sampling terminology used in this Handbook. For fundamental terms and definitions related to measurement see VIM [16]. For general terminology in analytical chemistry see the Eurachem Guide [35].

Contents

NOTE: This page is intentionally left blank.

1 Scope

The aim of this Handbook is to explain, with detailed and practical examples, some of the methods available for the 1) estimation of sampling uncertainty for *a predefined sampling procedure* and 2) for estimation of measurement uncertainty of results using this sampling procedure and subsequent analysis. The examples provide assessments of whether the uncertainty obtained from a given sampling procedure is fit for purpose, i.e. fulfils predefined requirements. Furthermore, the Handbook gives suggestions for design of sampling validation and quality control.

Although the annexed examples are for a limited selection of materials, the methods are generally applicable to most sample types.

NOTE: This page is intentionally left blank.

2 Abbreviations and symbols

The following abbreviations, acronyms and symbols occur in this Handbook.

Abbreviations and acronyms

AMC	Analytical Methods Committee, part of UK Royal Society of Chemistry
ANOVA	Analysis of variance
BIPM	International Bureau of Weights and Measures
CEN	European Committee for Standardization
CRM	Certified Reference Material
FAPAS	trade name of body that organises international proficiency tests
GUM	Guide to the Expression of Uncertainty in Measurement
HPLC	High Performance Liquid Chromatography
ICP-AES	Inductively Coupled Plasma - Atomic Emission Spectrometry
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
IUPAC	International Union of Pure and Applied Chemistry
JCGM	Joint Committee for Guides in Metrology
LOD	Limit of Detection
LOQ	Limit of Quantification
NIFES	National Institute of Nutrition and Seafood Research
NIST	National Institute of Standards and Technology
PT	Proficiency Testing
QA	Quality Assurance
QC	Quality Control
RANOVA	Robust analysis of variance
XRF	X-Ray Fluorescence

Symbols

CV	Coefficient of variation – RSD in %
D	Absolute range from difference
D_{anal}	Absolute range from difference for analysis
D_{meas}	Absolute range from difference for measurement
$D_{(\bar{x})}$	Absolute difference between a value and a mean value
d	Relative range from difference
df	Degrees of freedom
n	Number of measurements
RSD	Relative standard deviation
R_w	Within-laboratory reproducibility or Intermediate precision
s	Standard deviation of measured values (x_i)
$s_{\log}$	Standard deviation of $\log(x_i)$
s^2	Variance
s_{anal}	Analytical repeatability
s_{meas}	Measurement repeatability
s_{samp}	Sampling repeatability
SS	Sum of Squares in ANOVA calculations
SS_{anal}	Sum of Squares within group – analysis
SS_{meas}	Sum of Squares between groups – measurement
u	Standard uncertainty
u_{anal}	Analytical standard uncertainty
u_{meas}	Measurement standard uncertainty
u_{samp}	Sampling standard uncertainty
U	Expanded measurement uncertainty of result at a level of confidence of approximately 95 %, $U = 2 \cdot u_c$
${}^{\text{F}}U$	Uncertainty factor at a level of confidence of approximately 95 %
V	Total variation between samples – used in variography
x_i	Measured value
$\bar{x}$	Arithmetic mean value (average)
x_{ijk}	Measured value from target or batch (i), sample (j) and split (k)

3 Sampling in the measurement process

In this section we describe the possible steps in a measurement process

A complete measurement process, starting with primary sampling and ending in the analytical determination is shown in Figure 1. There may be many or few intermediary steps, such as transportation and preservation of samples. Each step gives rise to a contribution to the measurement uncertainty.

The process steps of taking the sample(s) from the sampling target and performing the physical sample preparation (shaded boxes) are generally considered part of sampling and are carried out prior to delivering the sample at the door of the laboratory. A more extensive overview is given in the Eurachem guide [24].

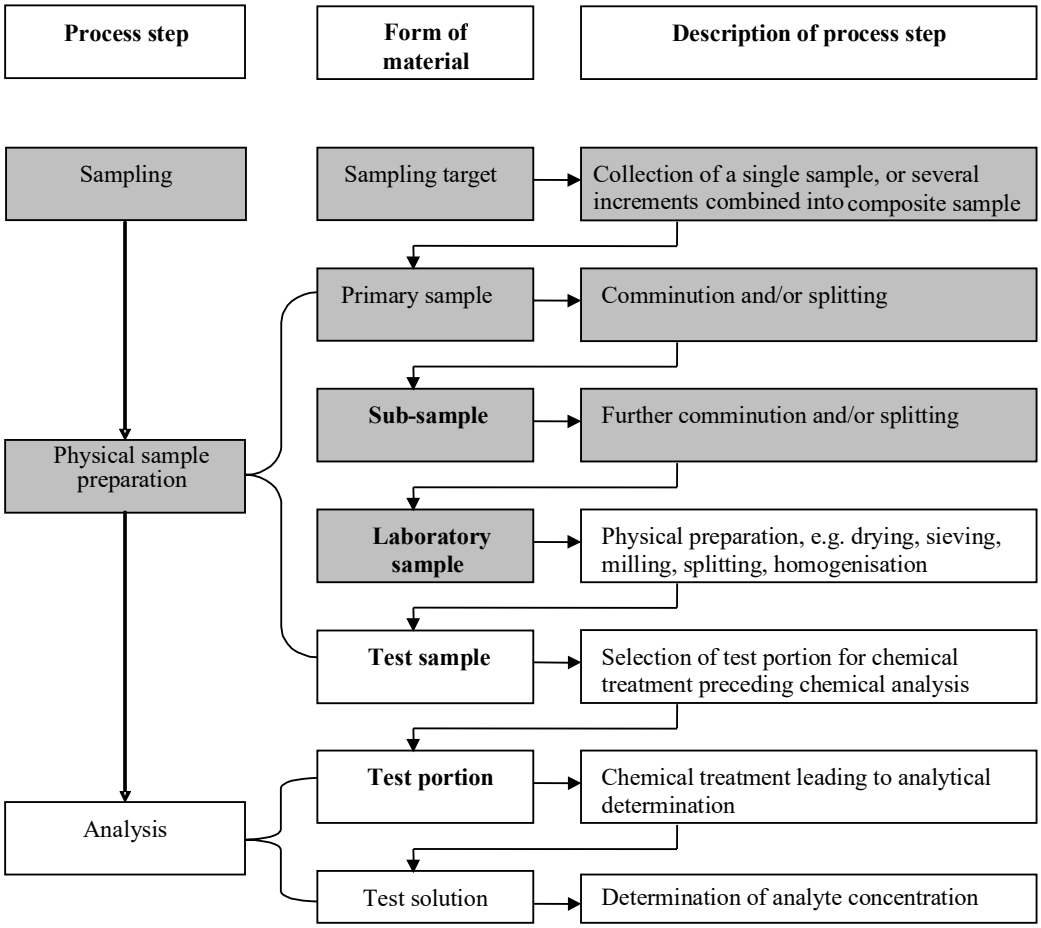


Figure 1 Schematic diagram of a typical measurement process including sampling, physical sample preparation (including transport) and analysis

NOTE: This page is intentionally left blank.

4 Purpose of sampling and quality requirements

In this section you will find guidance on how to include sampling uncertainty when planning your sampling programme

4.1 Purpose of sampling

The main purpose of most measurements is to enable decisions to be made. The credibility of these decisions depends on knowledge about the uncertainty of the measurement results. Uncertainty in measurement can be described as being made up of two components: 1) uncertainty derived from sampling a target, a mass of material (such as an area of land, or batch of food) and using those samples to represent the whole sampled target and 2) the uncertainty derived from the analytical process. If the uncertainty of measurements is underestimated, for example because the sampling is not taken into account, then erroneous decisions may be made that can have large financial, health and environmental consequences. For this reason it is essential that effective procedures are available for estimating the uncertainties arising from all parts of the measurement process. These must include uncertainties arising from any relevant sampling and physical preparation, as well as variability arising from material heterogeneity and chemical analysis – see further Section 7.

4.2 The sampling target

Based on the purpose of the measurement, one has to define the sampling target, i.e. what object is to be characterized (for example a produced batch of material, the soil of a contaminated site, etc.). It is important to properly define the sampling target; e.g. including also where and when do we want to measure. If there is time variation in the property measured, different sampling targets are possible, for example contaminant concentration at a factory outlet at the time of sampling, or the average outlet contaminant concentration over a year.

The definition of the sampling target becomes even more important when considering the uncertainty of the measurement. Still many measurement results are being presented to the end-user without any notion of the uncertainty, i.e. just as a number, x. However, the trend is to present the result as the measured value, x, with the associated expanded uncertainty, U, at a defined level of confidence given by the coverage factor. See also Section 5.4:

$$X = x \pm U \qquad \textbf{Equation 1}$$

The end-user will very naturally interpret that interval to be the concentration in the bulk material sampled; that is for the sampling target. In this view the uncertainty, U, includes any necessary allowance for heterogeneity in the bulk. The analyst, by contrast, might refer to the concentration and uncertainty in the sample received at the laboratory door, i.e. the laboratory sample. In metrological terms, this distinction arises because the two views are considering different measurands, i.e. quantities intended to be measured. One view is considering the concentration in the bulk material, or sampling target; the other the concentration in the laboratory sample.

These ambiguities in interpretation can only be avoided by careful specification of the measurand. The examples in this guide start with the specification of the measurand, which includes:

- sampling target – specification in space and time of the material to be characterized;
- parameter, including analyte where applicable – e.g. electrical conductivity or mass concentration of dissolved iron, Fe; and

- unit and base for reporting – e.g. weight-% reported on dry basis (105 °C, 2 h).

Here, it should be recalled that whereas the heterogeneity in time and/or space within the sampling target is contributing to the sampling uncertainty, it is often equally useful to have an estimate of the variation between targets, i.e. the variation due to heterogeneity outside the space and time defining the target. In the annexes the between-target variability is given as a CV in the summary table.

4.3 Quality requirements

Based on the purpose of the measurement it is necessary to set requirements for the required measurement quality, e.g. which uncertainty is acceptable. Examples of requirements are shown in Box 1.

Box 1 What kind of quality requirements could be useful?

The information we want can be *qualitative*:

- sampling from a batch of paint pots to determine whether, e.g. the colour of the paint matches the information on the label.

or *quantitative*:

- sampling of water in a wastewater stream to make sure that the concentration of a substance in the water does not exceed a permissible limit.

For *quantitative* information, we need to know how well the information describes the sampling target: The allowable concentration of a given substance is set to a certain maximum value. First of all we need to confirm if this is an average value over time (e.g. an annual mean concentration) or if it may never be exceeded, as this will affect the sampling frequency.

If the limit is given as a total emitted mass we also have to measure the flow of wastewater. Inhomogeneity in the wastewater has to be investigated and considered. If we take discrete samples at certain time intervals we also need to know if the concentration varies with time. This means that we need to consider, in quantitative terms, how well the sample taken describes the sampling target.

We will beforehand decide how well the samples taken must describe the target that we are studying: can we accept the quantitative information to have a 20 % uncertainty – thereby risking false alarms – or do we need a smaller uncertainty? Perhaps we can even accept a larger uncertainty, especially if we know that the measured level are well below the limit.

The sampling target we study is not homogenous and the properties vary, and there is an uncertainty associated with both sampling and analysis. The uncertainty of the measurement result interacts closely with decision making, Figure 2. Therefore, we always need to define with what certainty a decision shall be made and to control that the certainty of decision is attained through quantitative estimation of the measurement uncertainty including all steps in the chain.

Consequently, a definition of the measurement objectives (why) must always be done, the target must be defined (what, where and when), and the required decision certainty must be defined quantitatively. Based on this, quality requirements that are fit for purpose, *i.e.* are neither excessive (too expensive), nor insufficient (hopefully cheap), can be set.

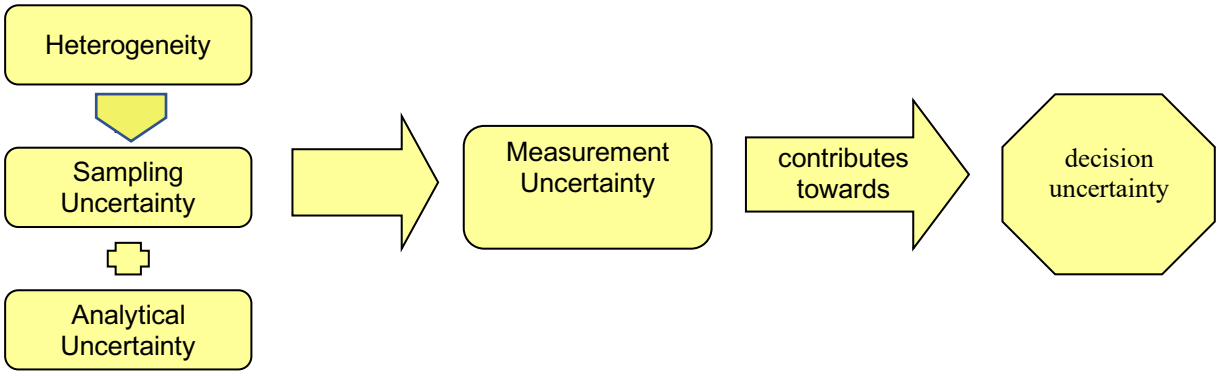


Figure 2 The uncertainty and decision chain – sampling (including heterogeneity) and analysis contribute to measurement uncertainty

Examples of measurement objectives are:

- control of limits, such as product specifications, environmental quality standards (EQS), maximum contaminant levels (MCL) or maximum residue limits (MRL);
- investigation of trends (variation in time) or distributions (variation in space).

The quality requirements are in many cases set as the analytical quality, e.g.:

- for the implementation of the EU Water Framework Directive: the limit of quantification (LOQ) must ≤ 30 % of the limit in the EQS and the expanded uncertainty must be ≤ 50 % of the limit [1];
- for control of tin in canned food in the EU: limit of detection (LOD) below 5 mg/kg, limit of quantification (LOQ) below 10 mg/kg, recovery 80 % - 105 %, and $HORRAT_R$ –value [2, 3] (requirement on between-laboratory precision) of less than 1.5 in a study of interlaboratory method performance [4];
- for environmental control in Denmark: maximum standard deviation at low concentrations, maximum CV and bias at higher concentrations, values set individually and for quality classes [5].

Conventionally, arbitrary quality requirements have been applied, for example:

- Limit of detection (LOD) below 10 % of limit, repeatability standard deviation better than 5 %, bias less than 20 % and all measurements within linear range [6].

Evidently, these requirements include analytical uncertainty only. If we want to ascertain that a decision can be made with a defined certainty, the basic requirement is that the measurement quality requirement (sampling + analysis) can be met. This means that we have to set quality requirements also for sampling. A complicating factor is that the uncertainty required to meet the measurement objective depends on the mean concentration and the limit to be enforced, see Box 2

Box 2 How can we set quality requirements from required certainty of decision?

The objective of a measurement was to decide whether a stockpile of soil was contaminated, i.e. it exceeded the maximum allowed contaminant level (MCL, here set to 100 units), and had to be disposed of. The required certainty of decision was 95 %. The target was the stockpile of soil.

The measurement must differ from the MCL by at the least 2 times[1] the standard uncertainty[2] in order to ascertain with a certainty of 95 % that the mean is different from the MCL, see Section 5.4.

The measurement result was at 80 units (80 % of the MCL), and a standard uncertainty of 10 units could thus be accepted while still attaining the required certainty of decision.

The analytical uncertainty was 4 units (CV 5 %), and that left room for a sampling uncertainty of 9.2 units (CV 11 %) while still maintaining the total uncertainty below the required value of 10 (CV 12 %).[3]

If the measurement result had been 50 units, a standard uncertainty of 25 units could be accepted and with the same relative analytical uncertainty (5 %), this would require a sampling uncertainty of not more than 24.9 units or almost 50 %. This would allow for taking significantly fewer sub-samples than with a measurement of 80 units while still maintaining the required decision certainty.

With this approach, it was possible to set quality requirements for the measurements considering the measurement objective, the required decision certainty and assumed contaminant concentration. The requirements could be set for the different steps in the measurement process and they could be set to reflect the acceptable uncertainty.

[1] In fact, a factor of 1.65 can be used if the aim is a 95 % one-sided statistical test.

[2] Uncertainty can be expressed as standard or expanded uncertainty – in this box uncertainty is discussed at the standard uncertainty level.

[3] $10 = \sqrt{9.2^2 + 4^2}$

5 Concepts of error and uncertainty in measurement

In this section we describe and discuss the different sources and nature of the errors that contribute to the total uncertainty of the measurements, including an introduction to how these may be estimated and expressed

5.1 Uncertainty sources

The uncertainty of a measurement result arises from a variety of sources, and these can be categorized in different ways. The uncertainty sources in the analysis step (see Figure 1) are well studied, but less focus has been given to those from sampling. In Table 1, some uncertainty sources in sampling and sample preparation are listed.

Table 1 Some uncertainty sources in sampling and sample preparation.

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

Often the sources believed to contribute to the uncertainty are presented in a *cause-and-effect-diagram*. A typical example is presented in Figure 3. An investigation of uncertainty sources can be a useful help in identifying those steps in the measurement process that could contribute to the total measurement uncertainty – see Box 3.

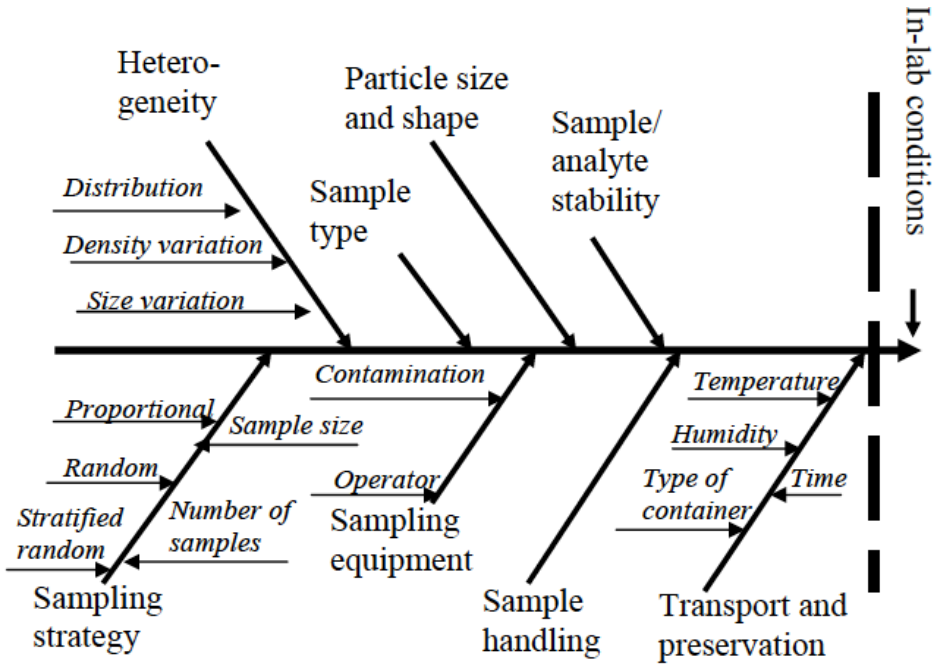


Figure 3 Cause-and-effect diagram of sources contributing to sampling uncertainty

Box 3 How can we use an investigation of the sources of uncertainty to pick the right point of action?

For a row of stockpiles of contaminated soil as described in Box 2 the sampling standard uncertainty, s_{samp}, turned out to be too high to meet the quality requirement. The measurement was 80 units, but the standard measurement uncertainty was 25 units, not 10 units as required. Using a replicate sampling design, Section 6, it was demonstrated that the excessive uncertainty was from the sampling (24.7 units), not from the analysis (4.0 units). A closer examination of the contaminant distribution in the stockpiles showed that contamination varied with depth because of depletion due to evaporation and leaching in the top 25 centimetres.

With the sampling originally done using a simple 50 cm core sampler, this heterogeneity resulted in highly variable samples taken from different positions in the stockpiles with different depths. Knowledge of the cause of the excessive measurement uncertainty made it possible to design a sampling protocol across a full cross section of a stockpile sampled for homogenization, splitting and subsampling that could provide the required lower measurement uncertainty of 10 units [7].

5.2 Systematic and random effects

The terms systematic effects (related to 'trueness') and random effects (related to 'precision') illustrated in Figure 4 are familiar to most readers from their work with estimating measurement uncertainty. The best accuracy (lowest measurement uncertainty) is achieved in case b) in Figure 4 where the individual results are all close to the reference value. In cases a) and b) there is no significant bias as the results are all clustered in the centre of the target. However, the precision is poorer in case a) as the results are more widely scattered. The precision in case d) is similar to that in case b). However, there is a significant bias in case d) as all the results are far away from the reference value. The accuracy is poorest in case c) as the results are widely scattered and are off-set to the right of the target.

The uncertainties caused by the sampling step can be divided into the same two categories, each being caused by a defined set of sources. Generally speaking, the systematic effects are hard to quantify but often possible to avoid, whereas the random effects are easier to quantify but harder to avoid. The methods for estimation of sampling uncertainty described in this Handbook generally quantify the precision only.

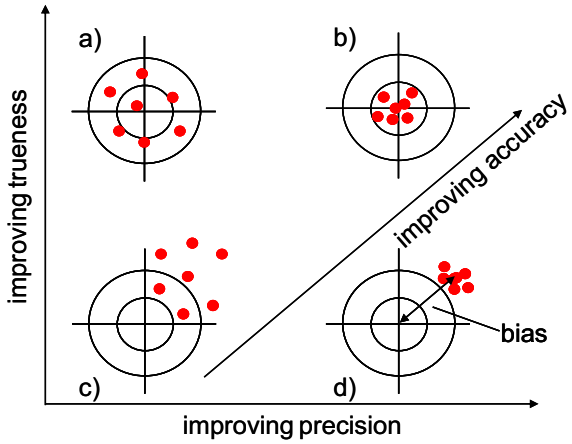


Figure 4 The 'shots' on the target represent individual measurement results; the reference value is the centre of the target, courtesy of Eurachem [35]

Systematic effects in sampling can be caused by the heterogeneity of the sampling target combined with an inability of the sampling procedure to properly reflect this heterogeneity. The heterogeneity can in turn be divided into the 1) *inherent*

heterogeneity of the material, caused by e.g. different size, shape and composition of the particles in a solid sample or different molecules in liquid samples, and 2) *distribution heterogeneity* caused by e.g. poor mixing, which may allow particles or molecules of different characteristics to segregate in the target. A very obvious example is particles in a stream of water that tend to fall downwards unless the stream is constantly and properly mixed, or two liquids that do not mix, e.g. oil in water. Systematic effects should always be accounted for in solid samples and particle-rich waters. In liquids the analyte may have to be stabilised after sampling to avoid systematic effects.

Even if the systematic effects are hard to quantify, as discussed above, there are some things we can do to reduce them:

- select methods for sampling and sample suitable for the sampling target and its properties (see also Box 4) such as e.g. grain size and size distribution, target heterogeneity, target layering, analyte instability etc;
- increase the size of the sample. It is obvious that if we sample and analyse the whole target, we will also get rid of the systematic effects. In almost all cases this is impossible and/or impractical, but increasing the sample size will give a better representation of the whole target;
- grinding solid materials. Reducing the particle size of either the whole target or taking out a relatively large sample, grinding it and then collecting a sub-sample, may decrease the systematic effects;
- mixing. This will reduce the segregation and can be applied both with solid samples and with liquid samples in e.g. a stream by selecting a sampling location where the stream is properly mixed. However, it should be noted that in some special cases mixing may induce the segregation. In these cases, mixing should be avoided;
- the composition of the sample caused by chemically and/or microbiologically induced changes during storage or transportation prior to the analysis.

Random effects are easier to quantify and can be minimized. They are mainly caused by variations in the composition of the sample in space or in time, variations that may be either cyclic or non-cyclic. Furthermore, random effects may be caused by variations in:

- the sampling protocol, e.g. if different protocols are used;
- sampling procedure or the handling of the sample, e.g. caused by different persons being involved;
- the sampling equipment and the way in which the equipment works.

The most obvious approach to reducing random effects is to increase the number of samples taken, which in turn will lead to a smaller standard deviation of the mean result. An equivalent approach is to increase the number of sub-samples or increments taken to produce one composite sample for investigation.

A careful investigation of the variations in time and space, carried out as part of the validation of the sampling procedure, might be needed to select the proper sampling frequency or spatial distribution for the given quality requirement. Collecting too many samples will just be more expensive, but will not necessarily give more or better information, and thus has to be avoided. Note that the suggested ways to reduce the systematic effects above will generally also decrease the random sampling effects. The effects and causes of systematic and random errors are further discussed in [8].

Box 4 How can knowledge of the type of sampling error help in designing the sampling?

The major source of uncertainty in contaminant measurements, for a row of soil stockpiles from Box 2, was shown to be varying contaminant concentration with depth, i.e. a systematic effect. Sampling or sub-sampling an increased number of replicates would not have reduced uncertainty because the effect of doing so is mostly limited for systematic errors. Instead, a more suitable procedure was designed, as described in Box 3.

5.3 Estimating sampling uncertainty

Both sampling and analysis contribute to measurement uncertainty. The uncertainty contribution due to physical sample preparation, transport, sample storage etc. should either be included in the sampling step or the analysis. The random part of the uncertainty is described by the standard deviation. The variance of the measurement, s^2_{meas} is given by the following equation:

$$s^2_{meas} = s^2_{samp} + s^2_{anal} \qquad \textbf{Equation 2}$$

The basic tool to estimate the size of the random part of the measurement uncertainty, s_{meas}, is to repeat the measurement, i.e. to sample the same target and analyse the samples.

The most practical way of estimating the sampling standard deviation, s_{samp}, is the double split replicate experimental design (see Section 6) where s_{anal}, can be obtained from variation between analytical replicates. The s_{samp} can then be obtained by rearranging Equation 2:

$$\boldsymbol{s_{samp} = \sqrt{s^2_{meas} - s^2_{anal}}} \qquad \textbf{Equation 3}$$

Box 5 Example of calculating the standard deviation for sampling

For a stockpile of contaminated soil described the analytical standard deviation was found to be 4 units and the measurement standard deviation to be 10 units in a double split replicate study. The sampling standard deviation can then be found using Equation 3:

$$\boldsymbol{s_{samp} = \sqrt{s^2_{meas} - s^2_{anal}}} = \sqrt{10^2 - 4^2} = 9.2$$

The systematic errors (bias) cannot be easily obtained, but some approaches to this are given in Table 2. As to the determination of the bias due to sampling – this is a more difficult task. Possible alternative approaches are:

- when a theoretical value is known, e.g. from production, and used as an estimate of the true value, see Annex C, or when sampling is performed on a reference sampling target [9];
- when two or more persons perform sampling and analysis, intersampler comparisons – e.g. when both producer and client perform sampling and analysis of the same batch of material, see Annex B;
- when comparing results with those obtained using a detailed reference sampling procedure such as e.g. for sampling coating powders [10];
- proficiency testing of sampling [37].

Table 2 Examples of tools for the estimation of uncertainty contributions from sampling

Random *(precision)*	***Systematic*** *(bias)*
Replicate samples	Reference sampling target Sampling proficiency testing schemes Intermethod comparisons Known value of sampling target Reference sampling procedure

5.4 Standard and expanded uncertainty

The standard uncertainty for measurement, u_{meas}, is calculated by combining the contribution from sampling, u_{samp} and analysis u_{anal}. As stated above in many cases the u_{samp} can be set equal to s_{samp} and u_{anal} can be obtained from the analytical laboratory. Then u_{meas} can be calculated:

$$u_{meas} = \sqrt{s_{samp}^2 + u_{anal}^2} \qquad \textbf{Equation 4}$$

The expanded uncertainty, U, of a single measurement, x, can be calculated from the standard uncertainty, u, applying a coverage factor of 2:

$$U = k \cdot u_{meas} = 2 \cdot u_{meas} \qquad \textbf{Equation 5}$$

This approach to reporting a measurement result with its associated uncertainty will give an interval having a level of confidence of approximately 95 %. The interval will thus include the "true value" with 95 % certainty. The uncertainty should ideally include all steps in the procedure from sampling to analysis and cover all important random and systematic effects.

NOTE: This page is intentionally left blank.

6 The replicate design for estimating sampling uncertainty

This section explains the most common designs of experiments where the contributions from sampling and analytical random errors to the total uncertainty can be estimated

The basic principle of the replicate design is to apply the same sampling procedure two or more times on the same target or on different targets. The replication can be performed with *single split design*, in particular if the analytical uncertainty is already well established since only s_{meas} is determined. With *double split design* other, critical, steps where information on the uncertainty is required can be investigated separately, as e.g. s_{samp} and s_{anal}. The double split design can be adapted/extended to provide information on other contributions, such as sub-sampling, preservation, transportation and storage of samples.

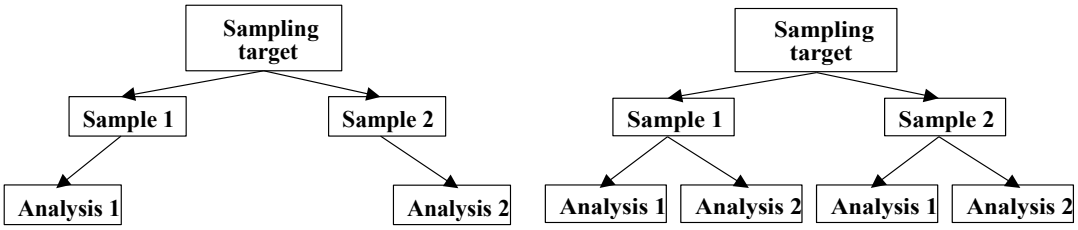


Figure 5 The principles of the replicate design with (left) single split and (right) balanced double split (two level nested design)

The replicate design is illustrated in Figure 5 for a design with *single split* and *balanced double split*. It should be noted that the double split might well be performed also with an *unbalanced design*, where only one of the two samples is analysed twice [24]. This might be a useful alternative if the sampling uncertainty is dominating and the analytical step is expensive or time consuming. The use of the symbols in the replicate design and calculation of mean range is shown in Table 3. At least eight replicates are needed to get a reliable estimate – the higher the number of replicates the better the estimated standard deviation will be.

Table 3 Symbols used in the replicate design showing the calculation of a mean range value for analysis, $\bar{D}_{anal}$ and for measurement, $\bar{D}_{meas}$

Rep #	Sample 1				Sample 2				
	x_{i11}	x_{i12}	$D_{i1} = \lvert x_{i11} - x_{i12} \rvert$	$\bar{x}_{i1}$	x_{i21}	x_{i22}	$D_{i2} = \lvert x_{i21} - x_{i22} \rvert$	$\bar{x}_{i2}$	$D_i = \lvert \bar{x}_{i1} - \bar{x}_{i2} \rvert$
1	x_{111}	x_{112}	$D_{11} = \lvert x_{111} - x_{112} \rvert$	$\bar{x}_{11}$	x_{121}	x_{122}	$D_{12} = \lvert x_{121} - x_{122} \rvert$	$\bar{x}_{12}$	$D_1 = \lvert \bar{x}_{11} - \bar{x}_{12} \rvert$
2	x_{211}	x_{212}	$D_{21} = \lvert x_{211} - x_{212} \rvert$	$\bar{x}_{21}$	x_{221}	x_{222}	$D_{22} = \lvert x_{221} - x_{222} \rvert$	$\bar{x}_{22}$	$D_2 = \lvert \bar{x}_{21} - \bar{x}_{22} \rvert$
3	x_{311}	x_{312}	$D_{31} = \lvert x_{311} - x_{312} \rvert$	$\bar{x}_{31}$	x_{321}	x_{322}	$D_{32} = \lvert x_{321} - x_{322} \rvert$	$\bar{x}_{32}$	$D_3 = \lvert \bar{x}_{31} - \bar{x}_{32} \rvert$
			$\bar{D}_{i1} = \frac{\sum D_{i1}}{n}$				$\bar{D}_{i2} = \frac{\sum D_{i2}}{n}$		$\bar{D}_{meas} = \frac{\sum D_i}{n}$
Mean range analysis			$\bar{D}_{anal} = \frac{\bar{D}_{i1} + \bar{D}_{i2}}{2}$						

NOTE: Replicate balanced double split design x_{ijk} for three sampling targets (i). For each replication of a target, two samples (j) are taken and from each sample is taken two test portions (k) for analysis.

7 Uncertainty contributions with replicate design

In this section we describe and discuss the contributions to the measurement uncertainty with the replicate design, and which contributions are included and neglected

The uncertainty contributions included in the calculation of measurement uncertainty using different replicate designs are further explained in Figure 6.

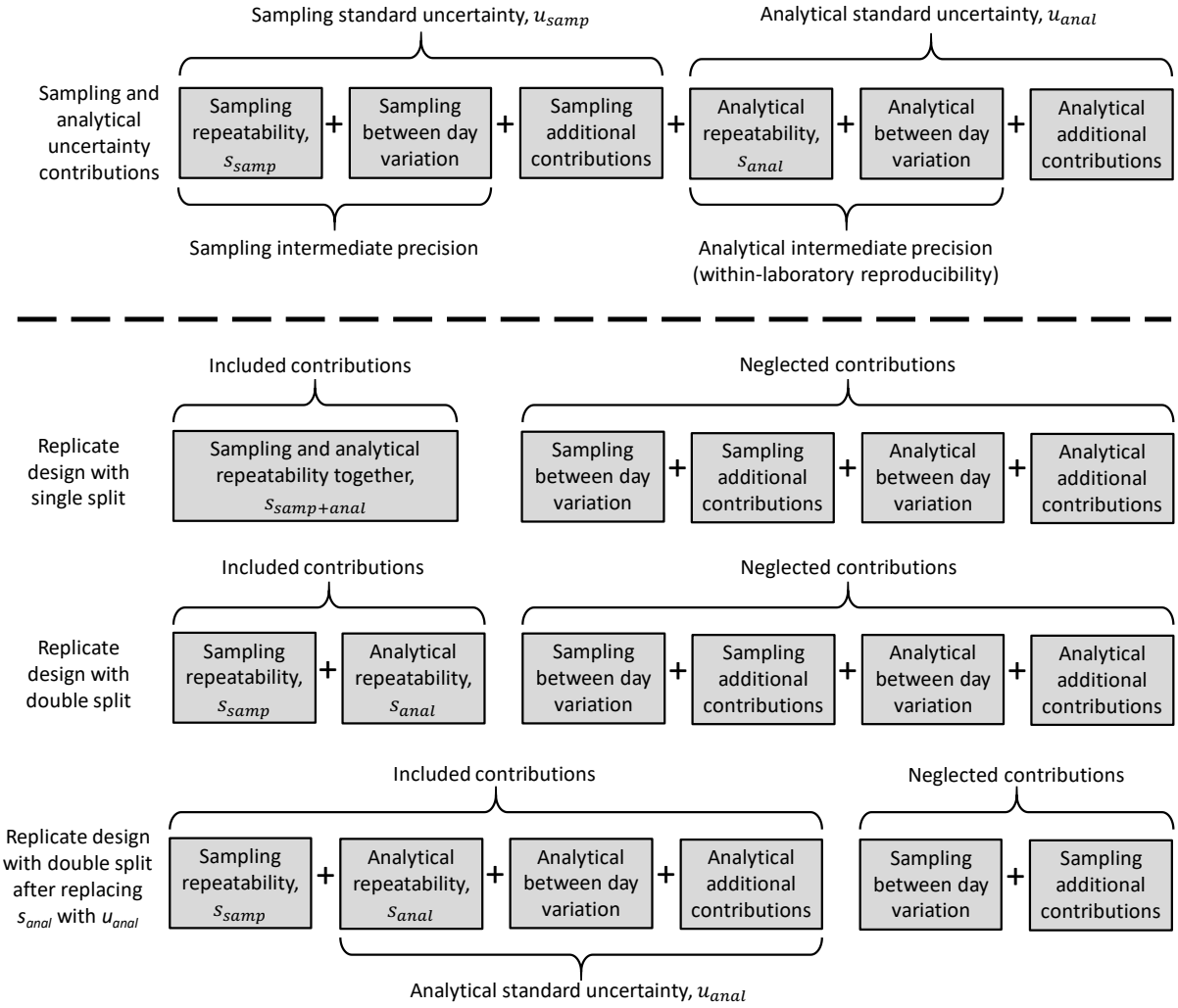


Figure 6 Illustration of the uncertainty contributions that are included using the replicate design with a single split at the sampling step, or a double split – split at the sampling step and split at the analytical step

The upper part in Figure 6 illustrates that the sampling uncertainty and the analytical uncertainty both consist of *repeatability*, *between day variation*[1], and *additional* contributions. The combination of repeatability and between day variation are often called intermediate precision. The intermediate precision for the analytical step is also called *within-laboratory reproducibility*, and this type of precision can be obtained from the analysis of one or several control samples used in internal quality control work [33]. Additional contributions often include handling of systematic effects but that is not

[1] The term "between day" is commonly used in the literature, however, here it would be more appropriate to call it "between occasion".

straightforward. This is further discussed in Section 5.2. The lower part of Figure 6 illustrates which contributions are included in the estimated uncertainties using different replicate designs.

Employing *replicate design with one split* at the sampling step (see Figure 5, left) will give the measurement repeatability, s_{meas}, which is the sampling and analytical repeatability together ($s_{samp+anal}$). Utilizing *replicate design with double split* at the sampling step and at the analytical step (see Figure 5, right) will give the sampling repeatability (s_{samp}) and analytical repeatability (s_{anal}) separately.

The uncertainty for the *neglected contributions* should be small, or the uncertainties need to be added separately. If not, the measurement uncertainty will be underestimated.

Especially for heterogeneous sampling targets, the sampling between day variation and the sampling additional contributions can in many cases be regarded as negligible compared to the sampling repeatability. Here, sampling between day variation can include e.g. variation caused by different samplers or different sampling equipment. However, *sampling additional contributions* can in some cases contribute considerably to the sampling uncertainty. Hence, the possibility to neglect some of the contributions should be considered case by case.

The analytical between day variation (including for instance variation in calibration of the instrument) and the *analytical additional contributions* are seldom negligible compared to the analytical repeatability. Hence, the analytical repeatability (s_{anal}) is not a good estimate of the analytical standard uncertainty (u_{anal}). In many cases it might be more appropriate to replace s_{anal} with u_{anal} that can be calculated from the expanded uncertainty for the analysis, U_{anal}, provided by the laboratory (normally $u_{anal} = U_{anal}/2$). However, U_{anal} from the laboratory can be based on larger variations in analytical conditions, and be valid for a larger scope (for example in covered sample matrices), than what is relevant here. If s_{anal} is found to be larger than u_{anal}, the heterogeneity of the analysed samples is likely larger than of the samples used when evaluating the analytical measurement uncertainty. In this case it might be better to use s_{anal}.

8 Principles of quality assurance in sampling

This section provides guidance on quality assurance of sampling, including the required competence, validation, and quality control of sampling procedures and their documentation

8.1 Competence requirements

To plan and perform qualified sampling and to make a reliable estimate of the measurement uncertainty the following competencies are required:

- competence about the issue and the sampling target – a specialist knowing the processes and variation in space and time. For blood sampling this would be trained medical staff, for sea-water sampling this would be a marine chemist/oceanographer, for production this would be a process engineer etc;
- theoretical and practical knowledge about the sampling procedure and the sampling equipment;
- competence about the sample from an analytical point of view e.g. stability, conservation, moisture uptake, how to avoid contamination and analyte loss etc;
- competence about the analytical method used, e.g. interferences, memory effects, sample amount needed, calibration strategy;
- competence about uncertainty in general.

In practical life, the responsibility for sampling may be with staff with analytical, technical or administrative background and the full suite of competencies will not be available to the person or even the institution in charge. Therefore, it is the obligation of the responsible person or institution to acquire the external competencies required to cover the entire field.

Box 6 How can the required competence be established?

The stockpiles of contaminated soils mentioned in Box 5 were situated at a soil remediation facility. A consultant was trusted with the task of making a risk assessment of the stockpiles before disposal. The consultant was a competent sampling planer. In order to supply the competencies required for sampling planning and uncertainty assessment, an engineer from the remediation plant (competence on the sampling target), a certified sampler (competence on sampling procedures, performance, quality control and documentation) and an analytical chemist from an accredited laboratory (competence on the sample treatment and analysis) were called upon.

The sampling competence may be sought with organisations or persons having their competences documented, e.g. by accreditation of the organisation to perform the sampling procedures [11] or certification of persons for environmental sampling [12].

8.2 Principles for sampling validation and quality control

Once the competence requirements have been set, the next step is to agree on the sampling and analytical uncertainty needed for the application ('target uncertainty') and choose an experimental design that enables the contributions to be evaluated effectively. To evaluate the sampling and analytical uncertainty two approaches (tools) can be chosen and combined: validation and continuous quality control.

Sampling validation comprises a one-time estimation determined under conditions expected in the routine use of the sampling procedure. The validation of sampling uncertainty may be done generically for the sampling procedure (initial validation) or specifically for the procedure used for the selected target (target validation). Initial validation is used when sampling is done as a one-time campaign (spot sampling,

example: contaminated site investigation) and target validation is done when sampling is done several times for the same target (repeated sampling, example: time or flow proportional sampling of wastewater). In effect, validation demonstrates what can be achieved and, if that conforms to the quality requirements, the procedures are deemed suitable for routine use.

Validation alone cannot ensure that routine results are indeed fit for purpose. Routine or target specific conditions may differ from those prevailing during the initial validation. This is especially true for sampling, where the larger part of the uncertainty component is often caused by the heterogeneity of the target. This is also true when a sampling procedure is applied at different targets. These circumstances emphasise the need for an ongoing quality control that includes sampling, to ensure that conditions prevailing at validation (and therefore the expected uncertainty attached to the results) are still applicable for every target and every time that the sampling and analytical procedures are executed. The combined use of validation and quality control is shown in Table 4.

Table 4 Illustration of the combined use of validation and quality control of sampling

	One procedure used for many sampling targets	***One procedure used repeatedly for one sampling target***
Validation	Initial validation yielding generic performance data for the procedure	Target validation yielding the performance data for the specific target and the procedure used
Quality control	Quality control with target specific verification of generic procedure performance data	Spot quality control verifying the performance data consistency over time

The need for internal quality control of sampling is not widely recognised at present, and methods for executing it are not well established, except in some specialised areas such as geochemical prospecting [13]. The methods used in validation are, with some simplification, applicable to quality control, but quality control is in most cases less extensive than validation. The reason for this is that validation needs to provide a good estimate of uncertainty, while quality control merely needs to demonstrate consistency over varying time and varying target compared to the uncertainty established at the validation.

The focus of quality control is almost exclusively the random aspect, whereas the systematic effects of sampling are difficult to address in validation and almost impossible in quality control. The flow in designing validation and quality control is shown in Box 7.

The principal tool for validation is replicate measurements, mostly in a split-level design; see Section 6 for description of the design. The validation must as a minimum provide the total (random) measurement uncertainty and a control of this against established quality requirement. In most cases, it is advisable to split the total uncertainty at least into a sampling and an analytical contribution. Additional splits can be useful, based on an analysis of the contributions to uncertainty from different sources. Annex C, Vitamin A in baby porridge, is used as an example of where the validation of a measurement process is designed in order to give information on the measurement uncertainty. In addition, the analysis evaluates a suspected point of high uncertainty, using a split replicate design.

The principal tool for quality control is also replicate measurements. This is minimally executed by taking two samples from the same target by a complete (and suitably randomised) duplication of the sampling procedure and with each sample analysed at least once.

Box 7 How can we design a validation and quality control programme for sampling?

The design of validation and quality control programmes is demonstrated in Annex A and Annex C for groundwater and baby porridge, respectively. An example of the basic steps is:

- acquire or determine the maximum measurement uncertainty acceptable from the required certainty of decision:
- analyse the measurement process and determine the expected points of high uncertainty;
- design and perform a validation study with at least 8 duplicate sampling events with separate analysis of each sample and with the sampling events varying in space (different points within the target) or in time (different sampling times) depending on the purpose of the sampling;
- include one or more split levels, if points of high uncertainty are anticipated;
- calculate the measurement uncertainty and the uncertainty associated with split levels, if pertinent;
- if the measurement uncertainty complies with the set quality requirement, accept the sampling procedure, design a quality control programme without split levels and construct a control chart for use in routine operation;
- if the measurement uncertainty exceeds the quality requirement, identify the critical point(s) of measurement and improve those;
- repeat the validation and confirm that quality requirements are now met.

In either case, continue routine sampling with the validated procedure and control the performance continually from the control charts and report the obtained measurement uncertainty to the customer.

The uncertainties can be calculated from quality control data as described in Section 9, and can be compared to the quality requirements and to the uncertainties obtained during validation. This approach requires not less than 8 sets of duplicate quality control results. If an early warning of a measurement process out of control is required, control charts can be used.

Range control charts[1] [14] are constructed from e.g. duplicate measurements where the difference between the results:

$$D = |x_1 - x_2| \qquad \textbf{Equation 6}$$

is calculated and the standard deviation of measurement

$$s_{meas} = {}^{D}/_{1.128} \qquad \textbf{Equation 7}$$

A one-sided range control chart can be constructed with a control limit of $2.83 \cdot s_{meas}$ (yellow/light grey, not exceeded in 95 % of control result) and an action limit of $3.69 \cdot s_{meas}$ (red/black, not exceeded in 99 % of control). For details on construction of the chart, see Box 8.

An out-of-control value *D* shows that the measurement may be associated with higher uncertainty than found during validation. Such a result is diagnostic and may stem from a disturbance in either sampling or analysis; the latter should be detected by standard methods of analytical quality control.

[1] Here is described absolute range charts. If the concentration range is wide between different target relative range charts are recommended.

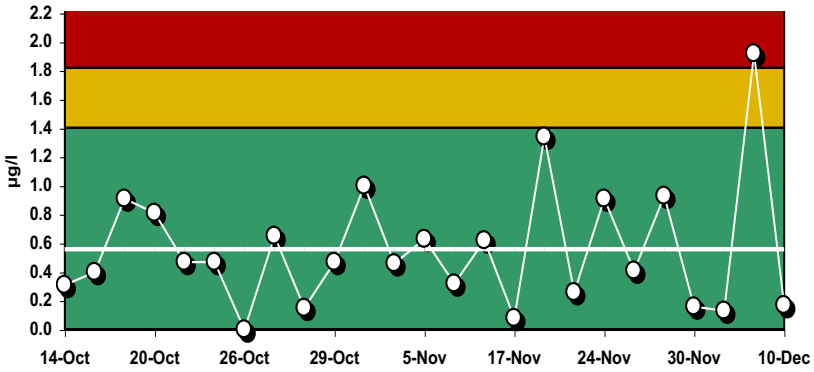


Figure 7 Example of a range control chart for quality control of measurement (sampling & analysis)

Box 8 How can a range control chart be constructed and used?

Construction of control charts is a routine task in analytical laboratories and is done by hand or by software packages. The basic steps are [14]:

- calculate the measurement standard deviation, s_{meas}, from the validation study;
- set the baseline to zero;
- set the central line, CL, to 1.128 · s_{meas};
- calculate the warning limit, WL, as 2.83 · s_{meas} and plot this value as a horizontal line in the chart;
- calculate the action limit, AL, as 3.69 · s_{meas} and plot this value as a horizontal line in the chart.

For each sampling occasion, perform at least one duplicate measurement (duplicate samples analysed separately)

- calculate the difference between the duplicate results as $D = |x_1 - x_2|$ and plot D in the chart;
- if D is above the action limit, do not report the result;
- if D is above the warning limit, check the two previous results. If one of those two is also outside the warning limit do not report the result.

Annex C demonstrates construction of a control chart for sampling of baby porridge. If the sampling is performed of different targets of varying concentrations of analytes, the same procedure is applied, but the relative standard deviation and the relative differences are used calculated as $d = \frac{|x_1 - x_2|}{\bar{x}}$.

8.3 Documentation

The documentation of sampling is needed in order to support all steps of the operations, from planning to evaluation of the measurements result. The different types of sampling documentation and their interrelation are described in Table 5.

Documentation should include at least:

- written sampling procedures;
- sampling field report;
- sampling reports.

Table 5 Summary of sampling documentation [14, 15]

Sampling method	A generic description of the operations used for sampling. The method may be a standard method, i.e. a method approved by a recognized, international standardization body.
Sampling procedure	A detailed description of the operations used for sampling according to a defined method and principle, and with defined equipment.
Sampling field report	The detailed notes on the sampling in the field.
Chain of custody report	A written record of the handling of the sample from sampling to analysis including transport and storage conditions.
Sampling report	Report summarizing the sampling including target definition, reference to applied method and procedure, relevant notes from field and chain of custody report.

Written procedures for how to take and handle the samples and for how to handle the sampling equipment are essential to ensure minimum variation caused by e.g. the involvement of different persons in the process. The sampling procedure is developed for each sampling organization and is normally developed from accepted or standardized sampling methods.

Sampling field reports serve to preserve the information of the details of the sampling process as observed during the sampling. Format of the sampling field report may vary in response to the need from one line in a procedure to an extensive report. The sampling report may be part of the measurement (analytical) report. Details on documentation of sampling can be found in [17].

NOTE: This page is intentionally left blank.

9 Calculations

In this section we present sampling uncertainty calculations. The contribution of random effects to the sampling uncertainty is based on replication by splitting the samples (replicate design) or replication in time or space (variography)

9.1 Calculation strategy

With the data available from performing the experimental design we need to choose:

- calculations either in concentration units or with log-transformed data;[1]
- method-range statistics, ANOVA, RANOVA or pooling standard deviations.

If data are in a small concentration range and the relative standard deviation, CV_{meas}, is less than about 15 %, the calculations can be performed using data in concentration units. Where the standard deviation is approximately proportional to the concentration and there is a wide concentration range calculations can be performed using relative range or on log-transformed data. When the standard deviation is greater than about 15 %, calculations should preferably be performed after log-transformation of the data – see example in section 9.6.

Selection of calculation method

The selection of the most appropriate method for statistical calculations will depend on a range of factors such as:

- the analyst's familiarity with different statistical methods;
- the complexity of the design behind the data; and
- the access to calculation tools such as spreadsheets and software packages.

As a help in selecting the appropriate method, the results obtained using four calculation methods, range (single and double split), ANOVA, and robust ANOVA for the same dataset (Annex C) are shown in Table 6. Robust ANOVA calculations were done with the software package RANOVA [21].

Table 6 Examples of results calculated using absolute range statistics, ANOVA and robust ANOVA on data in Annex C – vitamin A in baby food

	s_{anal} µg/100 g	CV_{anal} %	s_{samp} µg/100 g	CV_{samp} %	s_{meas} µg/100 g
Range – single split	-	-	-	-	42
Range – double split	30	8.6	19	5.5	35
ANOVA – double split	29	8.3	17	5.0	34
Robust ANOVA – double split	31	8.8	21	6.1	37

Evidently, the differences in statistical estimates obtained with the four different calculation methods are marginal in this case. It should, though, be emphasized that this need not always be the case, in particular with datasets with extreme outliers and skewed distributions.

[1] In this Guide common logarithms, base 10 (log) are used. The transformation can also be performed using natural logarithms (ln) as presented in the Eurachem Guide [24].

9.2 Range statistics

Range statistics are used to calculate the standard deviation, *s* or CV, for the steps in the split design. Normal distribution of the data is assumed – CV less than 15 %. The calculations can be done either by range or by relative range. Both may be used with single or double split designs.

- Range. The calculations are done from the difference between duplicate measurements. The uncertainties can be calculated if the analyte concentration does not vary with sampling position in time or space, and constant standard deviation over the measuring interval can be assumed.
- Relative range. The CV is calculated from the relative difference between duplicate measurements. This method is used when the analyte concentration will vary with sampling position (in time or space) and CV is constant over the measuring range[1]. Based on duplicate data, this has been suggested to be the case for most environmental and geochemical purposes at least with concentrations above ≈10 times the limit of quantification [18].

The calculation of the standard deviation from the mean differences is based on a statistical analysis of the relation between standard deviation and differences, and the factor applied depends on the replication chosen, e.g. duplicate, triplicate [14]. Similar estimates would be obtained if the standard deviations were calculated for each set of duplicates and combined as variances.

9.3 Single split design and relative range statistics

The relative range calculations are done with measurements of duplicates (*j*) on several sampling targets (*i*). Each set of duplicates producing the measurements x_{i1} and x_{i2}.

The absolute value of the difference, D_i, is calculated for each set of duplicates:

$$D_i = |x_{i1} - x_{i2}| \qquad \textbf{Equation 8}$$

The mean[2], $\bar{x}_i$, of the 2 measurements in each duplicate is calculated according to:

$$\bar{x}_i = \frac{x_{i1} + x_{i2}}{2} \qquad \textbf{Equation 9}$$

The relative difference, d_i, is calculated from the difference, D_i, and the mean for each set of duplicates:

$$d_i = \frac{D_i}{\bar{x}_i} \qquad \textbf{Equation 10}$$

The mean relative difference, $\bar{d}$, of *n* sets of duplicates is calculated:

$$\bar{d} = \Sigma d_i / n \qquad \textbf{Equation 11}$$

The relative standard deviation, CV, for measurement is calculated using a statistical constant [19] of 1.128 (when analysing duplicates):

[1] Calculations can also be performed on log-transformed data as recommenced in Section 9.6 and in the Eurachem Guide [24].

[2] In the double split design the mean value is calculated from all results on the sampling target – see further Appendix B.

$$CV = \frac{\bar{d}}{1.128} \cdot 100\ \% \qquad \textbf{Equation 12}$$

The standard deviation, s, at a given concentration, x_0, can be estimated from:

$$s = \frac{CV \cdot x_0}{100} \qquad \textbf{Equation 13}$$

An example of the calculations is shown in detail in Box 9.

Box 9 ***Calculation example demonstrating the use of relative range statistics for calculating* CV *from duplicates (single split design)***

Duplicate measurements of total Cr in soil (mg/kg) were done for samples taken at 10 positions and the calculations were done as follows

$\mathbf{x}_{i1}$	$\mathbf{x}_{i2}$	$D_i = \lvert x_{i1} - x_{i2} \rvert$	$\bar{x}_i = (x_{i1} + x_{i2})$	$d_i = D_i / \bar{x}_i$
20	10	10	15	0.67
223	157	66	190	0.35
312	150	162	231	0.70
816	432	384	624	0.62
55	125	70	90	0.78
54	124	70	89	0.79
442	325	117	384	0.31
765	755	10	760	0.01
232	516	384	324	1.19
650	215	435	433	1.01

Mean relative range of measurement	CV of measurement	Standard deviation of measurement at a level of 200 mg/kg
$\bar{d} = \sum d_i / n = 0.64$	$CV_{meas} = \frac{\bar{d}}{1.128} \cdot 100 = 57\ \%$	$s_{meas} = CV\ \frac{200}{100} = 114$ mg/kg

The application of relative range statistics is demonstrated in Annex A for validation and quality control of groundwater sampling. Note that in Box 9 the estimated expanded uncertainty will be more than 100 % giving an unrealistic interval including zero. In such cases we recommend to instead use log-transformed data and calculate an uncertainty factor, ${}^{F}U$ – see further Section 9.6.

9.4 Double split design and range statistics

The single split replicate design can be refined by introducing one (or more) additional set(s) of replicates or split(s), for example by doing replicate analyses of each of the two samples obtained according to the simple replicate design using duplicate samples (Figure 5). Annex C shows the use of the two split level replicate design for estimation of sampling and analysis uncertainty for measurements of vitamin A in baby porridge.

The calculation of the standard deviation requires that all measurements be within a range where the standard deviation is approximately constant. In Box 10, the calculation of the standard deviation for the different split steps is shown with the data from Annex C.

Box 10 Calculation example demonstrating the use of range statistics for calculating *s* from duplicate samples and duplicate analyses (double split design)

Duplicate samples (40 g) were taken from 10 batches of baby porridge and analysed for vitamin A (µg/100 g) in duplicate and calculations done as follows (see Annex C for more details).
Mean value of all results is 347.9 µg/100 g.

Sample 1				Sample 2				
x_{i11}	x_{i12}	$D_{i1}=\lvert x_{i11}-x_{i12}\rvert$	$\bar{x}_{i1}$	x_{i21}	x_{i22}	$D_{i2}=\lvert x_{i21}-x_{i22}\rvert$	$\bar{x}_{i2}$	$D_i=\lvert\bar{x}_{i1}-\bar{x}_{i2}\rvert$
402	325	77	363.5	361	351	10	356	7.5
382	319	63	350.5	349	362	13	355.5	5
332	291	41	311.5	397	348	49	372.5	61
280	278	2	279	358	321	37	339.5	60.5
370	409	39	389.5	378	460	82	419	29.5
344	318	26	331	381	392	11	386.5	55.5
297	333	36	315	341	315	26	328	13
336	320	16	328	292	306	14	299	29
372	353	19	362.5	332	337	5	334.5	28
407	361	46	384	322	382	60	352	32
$\overline{D}_{i1}=\frac{\sum D_{i1}}{n}$		36.5		$\overline{D}_{i2}=\frac{\sum D_i}{n}$		30.7	$\overline{D}=\frac{\sum D_i}{n}$	32.1

Mean range of analysis

$$\overline{D}_{anal}=\frac{\overline{D}_{i1}+\overline{D}_{i2}}{2}=33.6$$

Standard deviation of analysis

$$s_{anal}=\frac{\overline{D}_{anal}}{1.128}=29.8\ \text{µg/100 g}$$

$$CV_{anal}=\frac{s_{anal}}{\bar{X}}\cdot 100=\frac{29.8}{347.9}\cdot 100=8.6\ \%$$

Mean range of measurement

$\overline{D}=32.1$

Standard deviation of measurement based on duplicate analysis

$$s_{meas}=\frac{\overline{D}_{meas}}{1.128}=28.5\text{µg/100 g}$$

Standard deviation of sampling $s_{samp}=\sqrt{s_{meas}^2-\left(\frac{s_{anal}}{\sqrt{2}}\right)^2}=19.1$ µg/100 g

Comment: Since the analyses are based on a mean of duplicates the standard deviation of analysis is divided by square root of 2 in the equation above to give the standard deviation of the mean.

$$CV_{samp}=\frac{s_{samp}}{\bar{x}}\cdot 100=\frac{19.1}{347.9}=5.5\ \%$$

9.5 Double split design and ANOVA

Using the double split replicate design the standard deviations can also be estimated by applying analysis of variances (ANOVA). The variance is defined as the square of the standard deviation s^2. The source of the variation considered in this design will be the *between analyses variance* and the *between sample variance*. The values of s_{anal}^2 and s_{samp}^2 are estimated. An example is shown in Annex C .

The ANOVA calculations can be performed using dedicated software; e.g. RANOVA [21]. Note that Microsoft Excel does not offer ANOVA for the double split replicate design (two level nested design). There is often a small proportion (<10 %) of outlying values in the frequency distributions of the analytical, within-sample and between-sample variability. This may require the use of some method of down-weighting the effect of the outlying values such as the use of robust statistical methods for example *robust ANOVA*. This gives a more reliable estimate of the variances of the underlying populations if the measurements do not follow a normal distribution and have a significant number of outliers. In the dedicated software RANOVA, ANOVA and robust ANOVA can be used [21].

In this section we will describe in detail how the ANOVA calculations are performed. It should be emphasized that ANOVA calculations are more complicated and more rigorous than range statistics and more detailed information may be required, see e.g. [20]. It should be recalled that a basic understanding of the fundamentals of the methods is required in order to appreciate and consider the limitations and restrictions in their use. ANOVA for an *unbalanced design* is treated in the Eurachem Guide [24].

First, the variance of analysis is estimated based on the difference from the mean value, not on the range as in the approach described in Section 9.2. Given a two level split replicate design with duplicate samples (S1 and S2) taken and two subsamples (A1 and A2) analysed from each sample, the first step is to calculate the mean values of the analyses of each of the two subsamples:

$$\bar{x}_{i1} = \frac{x_{i11} + x_{i12}}{2} \qquad \textbf{Equation 14}$$

Then, for each of the two samples the squared differences $D^2_{(\bar{x})}$between each analytical result, x_{ijk}, and the mean value, $\bar{x}_{ij}$**,** of the two analyses of each sample is calculated. In this design the mean value $\bar{x}_{ij}$ is based on two measurements x_{ij1} and x_{ij2}, therefore the differences, from the mean value to each measurement for the samples, are equal (example given for the first sample);

$$|x_{i11} - \bar{x}_{i1}| = \lceil x_{i12} - \bar{x}_{i1} \rceil = D_{i1(\bar{x})} \qquad \textbf{Equation 15}$$

The sum of squares of differences of each sample is calculated as:

$$D^2_{i1(\bar{x})} + D^2_{i1(\bar{x})} = 2D^2_{i1(\bar{x})} \qquad \textbf{Equation 16}$$

The sum of squares of differences within groups, SS_{anal}, is calculated by summation of the sum of the squares of all the samples:

$$SS_{anal} = 2\sum_{\boldsymbol{i=1}}^{\mathbf{10}} \left[D^2_{i1(\bar{x})} + D^2_{i2(\bar{x})}\right] \qquad \textbf{Equation 17}$$

The degrees of freedom, df_{anal}, is calculated from

$$df_{anal} = i \cdot j \cdot k - i \cdot j \qquad \textbf{Equation 18}$$

where (*i*) is number of batches analysed, (*j*) number of samples from each batch and (*k*) number of test samples analysed of each sample.

The variance of analysis is then calculated as:

$$s^2_{anal} = {}^{SS_{anal}}/_{df_{anal}} \qquad \textbf{Equation 19}$$

and finally, the standard deviation and CV of analysis is calculated from:

$$s_{anal} = \sqrt{s_{anal}^2} \qquad \textbf{Equation 20}$$

$$CV_{anal} = \frac{s_{anal}}{\bar{X}} \cdot 100\ \% \qquad \textbf{Equation 21}$$

where $\bar{X}$ is the mean of all results across all targets.

In Box 11, the calculations are demonstrated for the same raw data used in Section 9.4 and in Annex C.

Box 11 Demonstrating of the use of ANOVA for calculating s_{anal} from duplicate samples and duplicate analysis

Duplicate 40 g samples (S1 and S2) were taken from 10 batches of baby porridge and analysed for vitamin A (µg/100 g) in duplicate (A1 and A2) and calculations done as follows (see Annex C for more details):

S1A1	S1A2	S2A1	S2A2	S1	S2	S1	S2
x_{i11}	x_{i12}	x_{i21}	x_{i22}	$\bar{x}_{i1}$	$\bar{x}_{i2}$	$2 * D^2_{i1(\bar{x})}$	$2 * D^2_{i2(\bar{x})}$
402	325	361	351	363.5	356	2964.5	50
382	319	349	362	350.5	355.5	1984.5	84.5
332	291	397	348	311.5	372.5	840.5	1200.5
280	278	358	321	279	339.5	2	684.5
370	409	378	460	389.5	419	760.5	3362
344	318	381	392	331	386.5	338	60.5
297	333	341	315	315	328	648	338
336	320	292	306	328	299	128	98
372	353	332	337	362.5	334.5	180.5	12.5
407	361	322	382	384	352	1058	1800

$\bar{X} = 347.9$	$SS_{anal} = 2\sum_{i=1}^{10}\left[D^2_{i1(\bar{x})} + D^2_{i2(\bar{x})}\right] = 16\,595$
$df_{anal} = i \cdot j \cdot k - i \cdot j$ = (10·2·2) - (10·2) = 20	$s^2_{anal} = {SS_{anal}}/{df_{anal}}$ = 16 595/20 = 829.8
$s_{anal} = \sqrt{829.8} = 28.8$ µg/100 g	$CV_{anal} = \frac{s_{anal}}{\bar{X}} \cdot 100 = \frac{28.8}{347.9} \cdot 100 = 8.3\ \%$

NOTE: The degrees of freedom of analyses is calculated from Equation 18 $df_{anal} = i \cdot j \cdot k - i \cdot j$ where (*i*) is number of batches analysed, (*j*) the number of samples from each batch and (*k*) the number of test samples analysed of each sample.

Second, the variance of sampling is estimated. The mean value of each batch, *i*, (two samples – 4 analyses) is calculated

$$\bar{X}_i = \frac{\bar{x}_{i1} + \bar{x}_{i2}}{2} \qquad \textbf{Equation 22}$$

Taking into consideration that the mean value of the batch is calculated from two values, the differences from the mean value of the batch to the mean values for each sample are equal. Therefore the square of differences between the mean value of the batch and the mean value of each sample in the batch is calculated according to:

$$(D_{i(\bar{x})})^2 = \left(\bar{X}_i - \bar{x}_{i1}\right)^2 = \left(\bar{X}_i - \bar{x}_{i2}\right)^2 \qquad \textbf{Equation 23}$$

The sum of squares of measurement, SS_{meas} is calculated according to:

$$SS_{meas} = \sum_{i=1}^{10} 4D_{i(\bar{x})}^2 \qquad \textbf{Equation 24}$$

The degree of freedom is calculated from the number of batches analysed, *i*, and the number of samples analysed of each batch, *j*:

$$df_{meas} = i \cdot j - i \qquad \textbf{Equation 25}$$

The variance of sampling is then calculated according to:

$$s_{samp}^2 = (SS_{meas}/df_{meas} - SS_{anal}/df_{anal})/2 \qquad \textbf{Equation 26}$$

The standard deviation, s_{samp}, and coefficient of variation, CV_{samp}, of sampling are calculated

$$s_{samp} = \sqrt{s_{samp}^2} \qquad \textbf{Equation 27}$$

$$CV_{samp} = \frac{s_{samp}}{\bar{x}} \cdot 100\ \% \qquad \textbf{Equation 28}$$

NOTE: If $s_{samp}^2 < 0$ then s_{samp} is conventionally set to zero.

In Box 12 the use of ANOVA for calculating standard deviation of sampling using the same raw data as in Box 11.

Box 12 Example demonstrating the use of ANOVA for calculating s_{samp} from duplicate samples and duplicate analyses

Duplicate samples (S1 and S2) were taken from 10 batches (i) of baby porridge and analysed for vitamin A (µg/100 g) in duplicate (A1 and A2) and calculations done as follows (see Annex C for more details):

S1A1	S1A2	S2A1	S2A2	S1	S2		
x_{i11}	x_{i12}	x_{i21}	x_{i22}	$\bar{x}_{i1}$	$\bar{x}_{i2}$	$\bar{x}_i$	$D^2_{i(\bar{x})}$
402	325	361	351	363.5	356	359.8	14.1
382	319	349	362	350.5	355.5	353	6.3
332	291	397	348	311.5	372.5	342	930.3
280	278	358	321	279	339.5	309.3	915.1
370	409	378	460	389.5	419	404.3	217.6
344	318	381	392	331	386.5	358.8	770.1
297	333	341	315	315	328	321.5	42.3
336	320	292	306	328	299	313.5	210.3
372	353	332	337	362.5	334.5	348.5	196
407	361	322	382	384	352	368	256

$\bar{x} = 347.9$	$SS_{meas} = \sum_{1}^{10} 4\, D_i^2 = 14\,231$	Data from Box 11 SS_{anal} = 16 595 df_{anal} = 20

$df_{meas} = (i \cdot j\text{-}i) = (10 \cdot 2\text{-}10) = 10$	$s^2_{samp} = (SS_{meas}/df_{meas} - SS_{anal}/df_{anal})/2 =$ $= (14\,231/10 - 16\,595/20)/2 = 296.7$
$s_{samp} = \sqrt{296.7} = 17.22$ µg/100 g	$CV_{samp} = \frac{s_{samp}}{\bar{x}} \cdot 100 = \frac{17.2}{347.9} = 5.0\ \%$

The output using the excellent software RANOVA is shown in Table 7 (classical ANOVA). The same result is obtained for s_{anal} and s_{samp} with classical ANOVA using manual calculations shown above as with RANOVA software. Similar results are obtained with robust ANOVA (Table 8) indicating that the data set has few outliers.

Table 7 Output from RANOVA software – classical ANOVA with input data from Box 11 – details see [21]

Mean	*347.85*		No. Targets	*10*
Total Sdev	*39.733*			
	Btn Target	**Sampling**	**Analysis**	Measure
Standard deviation	*21.268*	***17.224***	***28.805***	*33.562*
% of total variance	*28.65*	*18.79*	*52.56*	*71.35*
Expanded relative uncertainty (95 %)		*9.90*	*16.56*	*19.30*

Table 8 Output from using RANOVA software – robust ANOVA with input data from Box 11– details see [21]

Mean	*346.02*			
Total Sdev	*41.313*			
	Btn Target	**Sampling**	**Analysis**	Measure
Standard deviation	*18.137*	***21.218***	***30.456***	*37.119*
% of total variance	*19.27*	*26.38*	*54.35*	*80.73*
Expanded relative uncertainty (95%)		*12.26*	*17.60*	*21.45*

9.6 Log-transformed data and ^{F}U

If there is a wide concentration range or a standard uncertainty over 15 % the calculations are preferably performed on log-transformed data as shown in Box 13 using the same data as given in Box 9. These calculations will give an asymmetrical uncertainty interval that can be described by an uncertainty factor ^{F}U [24].

Box 13 ***Calculation example demonstrating the use of log scale calculations from duplicates (single split design)***

Duplicate measurements of total Cr in soil (mg/kg) were performed for samples taken at 10 positions and the calculations were done as follows:

$\mathbf{x}_{i1}$	$\mathbf{x}_{i2}$	Log $\mathbf{x}_{i1}$	Log $\mathbf{x}_{i2}$	s_{log_i}	$s^2_{log_i}$
20	10	1.30	1.00	0.21	0.05
223	157	2.35	2.20	0.11	0.01
312	150	2.49	2.18	0.22	0.05
816	432	2.91	2.64	0.20	0.04
55	125	1.74	2.10	0.25	0.06
54	124	1.73	2.09	0.26	0.07
442	325	2.65	2.51	0.09	0.01
765	755	2.88	2.88	0.00	0.00
132	516	2.37	2.71	0.42	0.18
650	215	1.30	1.00	0.34	0.05

Log standard deviation (pooled) of measurement	Uncertainty factor ^{F}U	Expanded uncertainty interval at a level of 200 mg/kg
$s_{log} = \sqrt{\sum s^2_{log_i} / n} = 0.24$	$^{F}U = 10^{2 s_{log}} = 3.0$	$67 - 600$ mg/kg

The uncertainty factor is calculated according to Equation 29:

$$^{F}\mathrm{U} = 10^{2 s_{log}} \qquad \textbf{Equation 29}$$

The upper confidence limit for the expanded uncertainty interval is calculated by multiplying the measured concentration by ^{F}U, and the lower confidence limit by dividing by ^{F}U. This confidence interval is therefore not symmetrical about the measurement value. In Table 9 is shown how the limits are calculated using an uncertainty factor giving an asymmetrical interval and using a relative uncertainty giving a symmetrical interval. The relative range calculation is shown in Section 9.2.

Table 9 Confidence limits (lower LCL, and upper UCL) calculated for a concentration of 200 mg/kg, when the measurement uncertainty is expressed as a relative expanded uncertainty and as an expanded uncertainty factor

Uncertainty		LCL mg/kg	Calculation of LCL	UCL mg/kg	Calculation of UCL
Calculations	**Value**				
Relative range	114 %	0 (-28)	200 - 114 %	428	200 + 114 %
Factor – data log-transformed	3.0	67	200/3.0	600	200 · 3.0
NOTE: Data from Box 9 and Box 13.					

9.7 Variography

Variography is used to determine variations in concentration in time or space within the sampling target. Here is described a procedure for time variation. e.g. where the samples are taken at the same spot in a flowing stream. Variography, together with knowledge about the analytical repeatability, s_{anal}, is a tool for identifying and quantifying (part of) the sampling repeatability, s_{samp}, components mainly caused by variations in time (but could also be in space).

In variography, a key tool is the variogram which is a plot of the variation between sample measurements taken at certain time intervals, e.g. 1 minute apart, 2 hours apart or x hours apart, against the time difference between the said samples. For the purpose of learning more about the nature of the investigated sampling target, the plot is useful to identify cyclic variations in time (or space), as the variability between samples taken will be lower if they are in the same period of a cycle. For the purpose of uncertainty calculations, the most useful feature of a variographic experiment is the ability to estimate the variability between two samples taken with zero time difference (called the 'nugget' intercept) by extrapolating the results to zero time difference. This estimate corresponds to the variability caused by material heterogeneity and the sampling process itself.

The sampling carried out to produce the time series data for a variographic analysis is called a variographic experiment. For this purpose, it is essential that the time elapsed between the taking the samples, the lag, is equal. Hence data from time-proportional samplings of e.g. a stream of wastewater in a discharge are very suitable for a variographic analysis. The variographic experiment can be carried out by using a time proportional automatic sampling equipment to take one sample per hour during 24 hours (to get an overview of the diurnal variation) and also to take as many samples as possible with the shortest possible time interval to investigate uncertainty from material heterogeneity and the sampling process, with the least possible interference from cycles or trends in the concentration of the material under investigation.

The variographic analysis and the interpretation of the results is best explained in an example, where we assume that the concentration of a certain parameter has been measured over a certain time period, with the resulting concentrations over time shown in Figure 8.

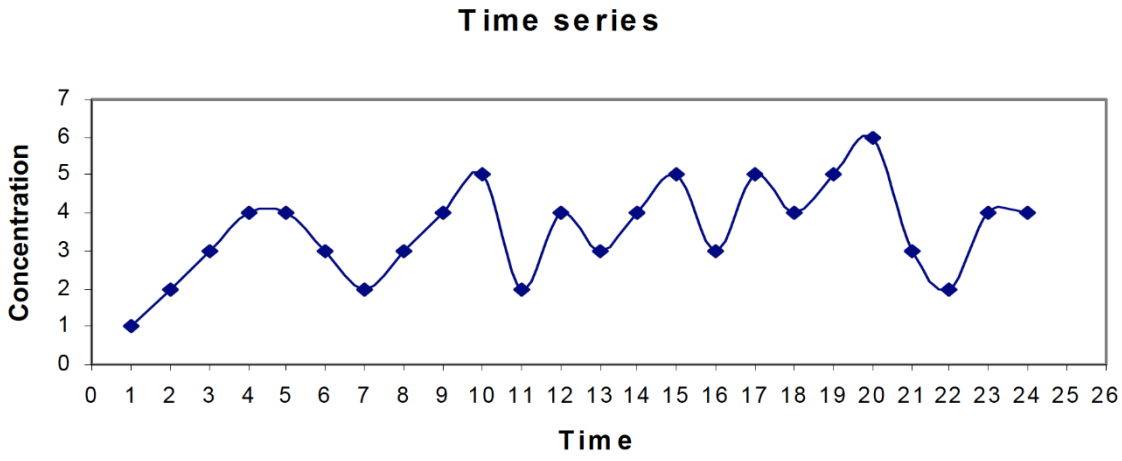


Figure 8 Example of a time series where the measured concentration is plotted against time (h) [23]

To construct the variogram, which is a calculation of the total variation, V, between the samples, separated by a constant time difference, e.g. 2 or 3 intervals, we calculate and plot the variations against the time intervals. The calculation of each point in the variogram is carried out as:

$$V(j) = \frac{\left[\sum_{i=1}^{n-j} (x_{i+j} - x_i)^2\right]}{\left[2 \cdot (n-j) \cdot (\bar{x})^2\right]} \qquad \text{Equation 30}$$

where x_i are the measured concentrations at the times i, j is the lag (time interval) between the results for which the variance is calculated, n is the total number of measurements in the time series and $\bar{x}$ the mean concentration over the time series. A variogram based on the time series given in Figure 8 is shown in Figure 9.

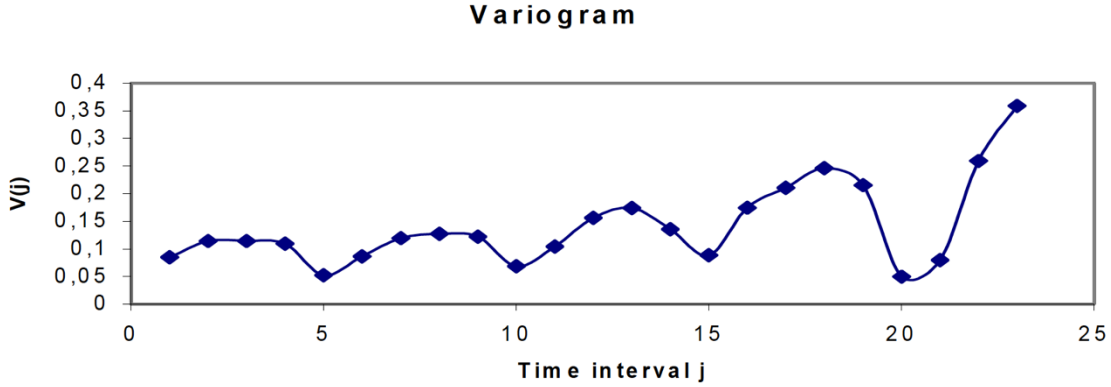


Figure 9 Variogram for the time series in Figure 8. On the *x*-axis are given the different time intervals, j, used to calculate the variances, *V*(j) on the *y*-axis

From the variogram it is obvious that the process in question has a periodic cycle consisting of 5 h intervals. This is virtually impossible to see from the plot of the time series and provides valuable information when designing the sampling procedure (helps to reduce systematic sampling errors), i.e.: variograms sometimes reveal variations that are not possible to see from the original time series. When designing the sampling procedure it is crucial for the choice of, for instance, sampling intervals to have

information about cyclic variations of the sampling target. Note that this means that the points to the right in the Figure 9 are based on very few measurements and thus are increasingly uncertain. In fact, the degrees of freedom for *V(23)* becomes zero, as it is based on one difference only (the 24 hour measurement minus the 1 hour measurement). In many cases the variances in the last 10 points of the diagram should therefore be neglected or at least interpreted very cautiously.

If the variogram is extrapolated to the *y*-axis the resulting value, *V(0)*, represents the minimum variation between two samples taken at closer and closer intervals using the sampling procedure in question. This minimum variation thus represents the variation caused by the actual sampling (material heterogeneity, variations in sampling), and variations caused by the analysis. The minimum variation thus quantifies how much of the total variation that comes from the measurement itself: sampling, sample treatment and analysis. From *V(0)* it is possible to estimate the standard deviation, *s*(0), representing material heterogeneity, variations in the sampling process and variations caused by the analysis:

$$s^2(0) = V(0) \cdot \bar{x}^2 \qquad \textbf{Equation 31}$$

then recalculated to a coefficient of variation, CV:

$$CV = \frac{s(0)}{\bar{x}} \cdot 100 \qquad \textbf{Equation 32}$$

s(0) is a representation of s_{meas} without heterogeneity caused by fluctuations in the process (representativity). The possible variations caused by differences in sampling equipment and operator, as well as variations from repeated setting up of the sampling equipment are not included. However, in properly conducted sampling this source of variation is small. To get the best possible estimation of *V(0)* it is important to perform the variographic experiment with a high sampling frequency, in order to be able to perform the best possible extrapolation to the *y*-axis. For strongly cyclic processes, a minimum in the variogram might be a better representation of *V(0)* than an extrapolation to the *y*-axis, in particular if the variographic experiment close to the *y*-axis is performed in an unfavourable part of the cycle. If this is the case, the extrapolation might easily overestimate *V(0)*.

The mathematical minimum variation *V(0)* is always positive and is often called the nugget effect. In the point *V(0)*, the process variation is neglected, and the point will thus as said above, in case the flow can be considered constant, represent the uncertainty sources that are caused by sampling and analysis, excepting representativity and possible variations caused by differences in equipment etc. These two components are independent, and the total variation is described in Equation 2: $s_{meas}^2 = s_{samp}^2 + s_{anal}^2$

V(0) will represent the s_{meas}. By inserting information about the analytical uncertainty, e.g. from laboratory quality control, the uncertainty from sampling may be estimated as described in Equation 3: $s_{samp} = \sqrt{s_{meas}^2 - s_{anal}^2}$

In case of a stream of wastewater, the model has to be extended to include also the variations caused by the measurement of the flow and the process variation. The model must then cover the following sources of variation:

- sampling, including material heterogeneity;
- sample handling and analysis;
- flow measurements;
- process variation.

The s_{samp} thus determined gives the sampling uncertainty in a single sample and it shows the effect of material heterogeneity and the sampling process. It gives a clear indication of the uncertainty caused by lack of mixing in the place where the sample is taken and is therefore a useful tool to estimate the suitability of the sampling arrangements. The variographic approach is demonstrated in detail in Annex D

NOTE: This page is intentionally left blank.

Annex A Groundwater

Measurand				Uncertainty estimation		
Analyte & technique	**Unit**	**Sector & matrix**	**Sampling target**	**Purpose**	**Design**	**Statistics**
Dissolved iron, ICP-AES	mg/L	Environment groundwater	The groundwater near one selected monitoring well	Measurement	Balanced double split	Relative range

A1 Scope

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

A2 Scenario and sampling target

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

The groundwater body is in glacial outwash sand with Miocene sands and clays below and glacial till above. The geology at the site is complicated with several local aquifers and aquitards.[1] The groundwater body as identified is 2 km × 2 km × 10 m, situated 20-30 m below the surface. The natural quality of the groundwater is anaerobic without nitrate, with sulphate and reduced iron, but without hydrogen sulphide and methane. One of the threats to the groundwater body is oxygen intrusion into the aquifer as the result of the water abstraction and concomitant groundwater table draw down.

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

To ensure the compliance of the monitoring program with the stated objective, a sampling validation study was initially conducted including all wells available and based on the results from this, a routine sampling quality control program was set up for implementation with the monitoring program for the selected monitoring well.

The properties of the groundwater body were summarised based on previous monitoring activities (surveillance monitoring). A summary for the two key parameters is shown in

[1] Aquifer: underground layer of water-bearing permeable rock, or permeable mixtures of unconsolidated materials. Aquitard: geological formation of layers comprised either of clay or on non-porous rock that restricts water flow from one aquifer to another

Table A 1. The standard deviation here includes variability in time and space (between targets) as well as measurement (sampling and analytical) uncertainty, i.e. it is the total variation.

Table A 1 Key chemical parameters for nine wells of the groundwater body, from surveillance monitoring

	Redox potential	Dissolved iron
	mV	mg/L
Mean	-123	1.1
CV[1]	27 %	56 %
Main cause of uncertainty	Oxygen impact during sampling and on-line measurement	Filtering of sample prior to analysis
[1]The CV includes variability in time, space (between targets as well as measurement uncertainty, i.e. it is the total variation.		

The chemical data suggest that the groundwater composition is quite uniform over time and space with respect to the main components (data not shown, CV 1.9-16 %), whereas the variability is high for the redox parameters (oxygen, redox potential and dissolved iron). The expected main causes of uncertainty are indicated in Table A 1 for the two key parameters and the causes were controlled during sampling.

A3 Sampling procedure

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

A4 Study design – double split replicates

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

Important is to choose between absolute or relative range calculations. Since the analyte concentration vary with sampling position a factor of 4 relative range calculations is performed – see further Section 9.2.

A4.1 Validation

The objective of the validation was to ensure that a measurement uncertainty, meeting the set quality requirement could be obtained, and to describe the components of

[1] The water table of a bottom sediment in a well is the upper surface of the zone of saturation. The zone of saturation is where the pores and fractures of the ground are saturated with water

uncertainty in order to identify points of improvement, if required. The validation programme was set up with sampling of six wells, two independent samplings per well and 2 sub-samples per sample analysed, see Figure A 1.

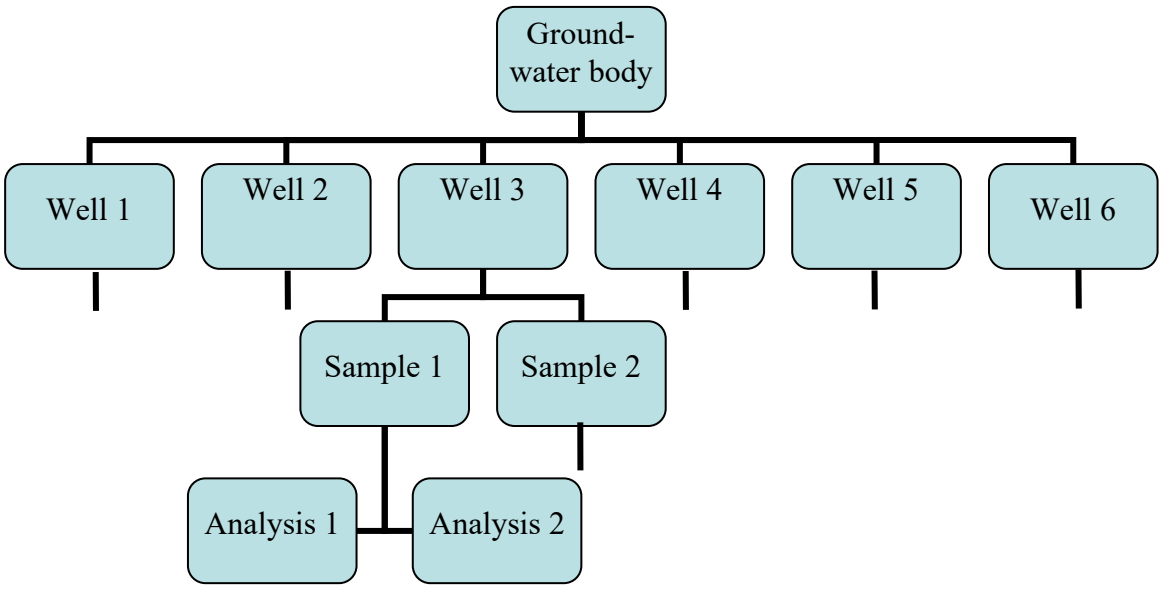


Figure A 1 Design outline for validation

A total of 12 samples were taken and 24 sub-samples were sent for analysis in one sampling round as validation study. No groundwater samples had measurement values of dissolved oxygen above 0.1 mg/L. The low redox potential measured (-200 to -110 mV) is consistent with the absence of oxygen (<0.1 mg/L) and the high dissolved iron concentrations (0.92 to 2.8 mg/L).

A4.2 Quality control

The objective of the quality control programme for the operational monitoring was to ensure that measurement uncertainty did not increase over time during the monitoring. The quality control programme was set up after careful evaluation of the results from the validation study. Quality control was designed to include duplicate sampling and each with duplicate analysis on one of the two annual sampling occasions of the monitoring programme, see Figure A 2. The quality control programme included six sampling occasions in one monitoring well.

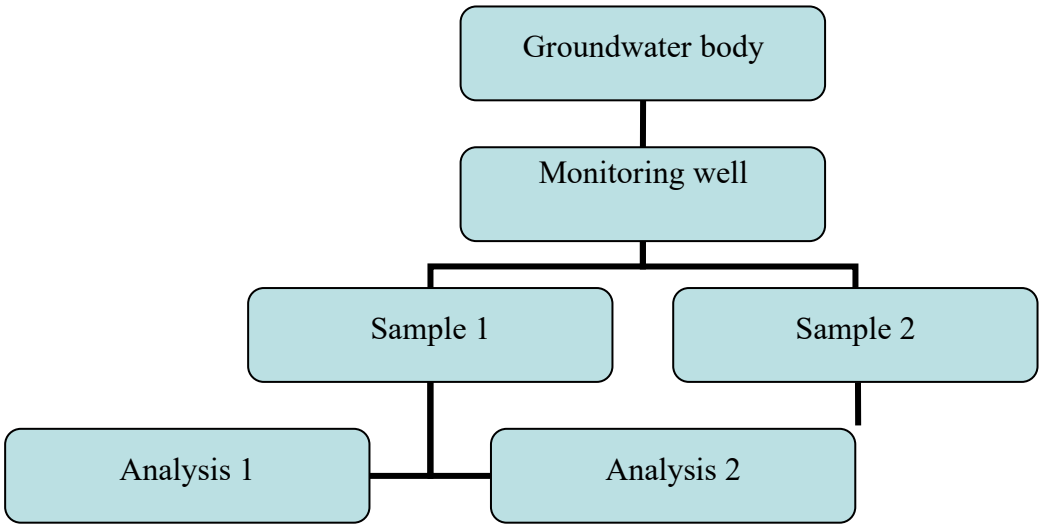


Figure A 2 Design outline for quality control, shown for one sampling occasion

The sample preparation and analytical set up for the two key parameters (redox potential and dissolved iron concentration) are shown in Table A 2.

Table A 2 Preparation and analytical programme

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

A5 Sample preparation and analysis

Duplicate online measurements/sub-samplings for laboratory analysis were done by taking out split sample streams and treating each stream independently. This means that the "analytical uncertainty" obtained with the duplicate design also included sub-sampling, pretreatment, such as filtering, and transportation.

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

A5.1 Field analysis

The sample stream was pumped through an on-line measuring array of a flow-through cell with sensors set up in series. The sensor used for redox potential is described in Table A 3. No quality control was performed of on-line measurements in the field.

Table A 3 On-line sensor used for redox potential measurements

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

A5.2 Laboratory analysis

Analyses were performed at an independent laboratory using an accredited method subject to the required quality assurance and analytical quality control. Method and performance data from analytical quality control are shown in Table A 4.

Table A 4 Method and performance data from quality control for laboratory analyses of iron

Technique	Repeatability	Within-lab reproducibility	Expanded uncertainty	Detection limit
ICP-AES	0.95 %	4.3 %	8.6 %	0.01 mg/L
NOTE: Performance data is for iron levels above 0.2 mg/L				

The reference material VKI Metal LL2, having a certified iron concentration of 0.200 mg/L, was used for quality control giving an estimate of analytical bias of +1.9 % from 92 control results.

The replicate data were treated using the relative range method, see Section 9.2. The applied calculations methods are demonstrated in Table A 5. For comparison, uncertainty estimates were calculated using ANOVA, see Section 9.3, and RANOVA [21, 22].

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

A6 Results

The relative range calculations in the validation study is shown in Table A 5 for dissolved iron. The calculations for redox potential was done similarly.

Table A 5 Relative range calculations for the validation study, dissolved iron

Well	S1A1 mg l^{-1}	S1A2 mg l^{-1}	S2A1 mg l^{-1}	S2A2 mg l^{-1}	Mean mg l^{-1}	d_1 %	d_2 %	d %
99.474	0.815	0.834	0.912	0.893	0.86	2.2	2.2	9.03
99.468	1.8	1.83	1.94	1.93	1.88	1.6	0.5	6.40
99.469	1.69	1.68	1.79	1.77	1.73	0.6	1.2	5.48
99.916	2.62	2.61	2.83	2.84	2.73	0.4	0.4	8.07
99.327	1.66	1.63	1.58	1.59	1.62	1.9	0.6	3.72
99.371	1.52	1.53	1.47	1.50	1.51	0.7	2.0	2.66
				Mean	1.72	1.21	1.14	5.89
				s	0.604			
	d_{anal} =	1.18				Analysis	CV_{anal} =	1.04 %
	d_{meas} =	5.89	CV_{meas} =	*5.22*		Sampling	CV_{samp} =	5.22 %
					Between target		CV_{target} =	35 %

The data from the validation study (6 different wells) using relative range calculations are shown in Table A 6.

Table A 6 Relative expanded uncertainty for analysis and sampling from validation data using relative range calculations

	Analysis	Sampling
Redox potential	5.2 %	15 %
Dissolved iron concentration	2.1 %	10 %

For comparison, results are shown in Table A 7 using absolute range, ANOVA and RANOVA (robust analysis of variance). The expanded uncertainty did not provide statistical estimates more than slightly different from those obtained with the simple range calculations. Also the absolute range calculations show similar results in this case where the concentration only varied a factor of four.

Table A 7 Relative expanded uncertainty for analysis and sampling for dissolved iron concentration from validation data using different calculations

	Analysis	Sampling
Relative range	2.1 %	10.4 %
Absolute range[1]	1.8 %	10.5 %
ANOVA	1.6 %	9.6 %
RANOVA	1.8 %	9.9 %
[1]Absolute range calculations results for dissolved iron from the Eurachem Guide [24].		

The results obtained with the range statistics during quality control (six sampling occasions) are shown in Table A 8.

Table A 8 Relative expanded uncertainty for analysis and sampling for quality control using relative range calculations

	Analysis	Sampling
Redox potential	18 %	3.8 %
Dissolved iron	2.5 %	3.6 %

In the quality control scheme of monitoring (data and calculations not shown), the variability between sampling occasions (between target, 9.9 %) was dominating the total uncertainty for parameters analysed as laboratory analysis (dissolved iron concentration, 2.5 % uncertainty), whereas the analytical uncertainty (18 %) was almost as important as the between target uncertainty (23 %) for on-line measurements (redox potential). The reason for the large contribution from on-line measurements is that during quality control, duplicate on-line measurements were done with two different instruments in contrast to the validation study done with one single instrument for both duplicate measurements. Accordingly, the analytical uncertainty including a contribution from instrument to instrument variation for redox potential was considerably larger in the quality control (18 %) than in the validation study (5.2 %). For dissolved iron concentration, the analytical expanded uncertainty was comparable in validation and in the subsequent quality control (2.1 % and 2.5 %, respectively). The expanded sampling uncertainty was lower when sampling just one well at different occasions during quality control (3.6 - 3.8 %) than when sampling different wells at the same time during validation (10 - 15 %). The CV for between target (variation from

one sampling occasion to the next) during quality control was small for dissolved iron concentration (9.9 %), but larger for redox potential (23 %).

If a continuous control of sampling uncertainty had been required, the control data could have been plotted in a range control chart, see Section 8.2, in order to obtain an early warning of excessive uncertainty (random errors) for each sampling occasion.

A7 Comments

The number of replicates in this study (6) was less than used in most cases and the risk of a decreased confidence in the uncertainty estimates should be considered in evaluation of the results.

The uncertainty contribution from sampling bias was only addressed through evaluation of the consistency of the measurements obtained from different, interrelated chemical parameters (concentrations of oxygen and dissolved iron, redox potential), and the evaluation supported that sampling and sample pretreatment had succeeded to avoid bias from oxygen impact and filter clogging.

A8 Assessment of fitness for purpose

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

A9 Reporting and interpretation

Single measurement data for dissolved iron concentration from the monitoring well shall be reported with an expanded, relative uncertainty of 4.0 %, as long as the monitoring quality control supports that this uncertainty is maintained.

A10 Summary

Dissolved iron concentration in groundwater	Expanded uncertainty			Between-target variability CV
	Sampling	Analysis	Measurement	
Validation	10 %	2.1 %	11 %	35 %[1]
Quality control	3.6 %	2.5 %	4.4 %	9.9 %[2]

[1]In the validation study, between-target variability was between wells
[2]In the quality control, between-target variability was between sampling occasions

NOTE: This page is intentionally left blank.

Annex B Iron ore

Measurand				Uncertainty estimation		
Analyte & technique	**Unit**	**Sector & matrix**	**Sampling target**	**Purpose**	**Design**	**Statistics**
Total iron XRF	% Fe (w/w) dried sample	Mining iron ore	Lot – 24 hours production	Sampling	Balanced triple split[1]	Range
[1]Triplicates from each lot (batch) analysed in duplicates						

B1 Scope

The scope is the determination of the sampling uncertainty of iron content in highly upgraded iron ore pellets produced at LKAB, when sampling is carried out according to the ISO standard 3082 for iron ores.[1]

B2 Scenario and Sampling Target

LKAB's main product is iron ore pellets. These are produced from finely ground, highly concentrated iron ore mixed with additives (one or more of dolomite, olivine, quartzite and limestone) and a binder before being rolled into 10 – 15 mm balls prior to oxidizing sintering at 1250 °C. The sampling target, the lot, is one day (three shifts, 24 h) of pellet production.

B3 Sampling procedure

The sampling shown in Figure B 1 follows ISO 3082 (2000) for iron ores. Sampling of the pellets in the pelletizing plant is realized by an automatic sampler from a conveyor belt. One increment for the *primary sample* is taken every 4 minutes, equivalent to approximately 300 kg per hour. After one hour, the sample (300 kg) is automatically divided by splitting. One part is used for screening analysis, and one part (approx. 30 g) is retained for chemical analysis. After 8 hours, these eight 30 g portions, the *sub-samples*, are mixed together into a 240 g sample and ground automatically. The three 240 g samples, the *laboratory sample* produced during one production day are transported to the analytical laboratory where they are mixed and split into a 150 g portions. One 150 g portion, the *test sample*, is dried, and from this dried *test sample* a *test portion* of 0.5 g is analysed. The number of increments from one lot in this case is 360 (every four minutes under 24h).

This sampling procedure is a general design for quantifying several parameters of iron ore pellets such as particle size distribution, metallurgical and mechanical properties.

1 ISO 3082:2000. Iron ores – Sampling and sample preparation procedures. (New version 2018)

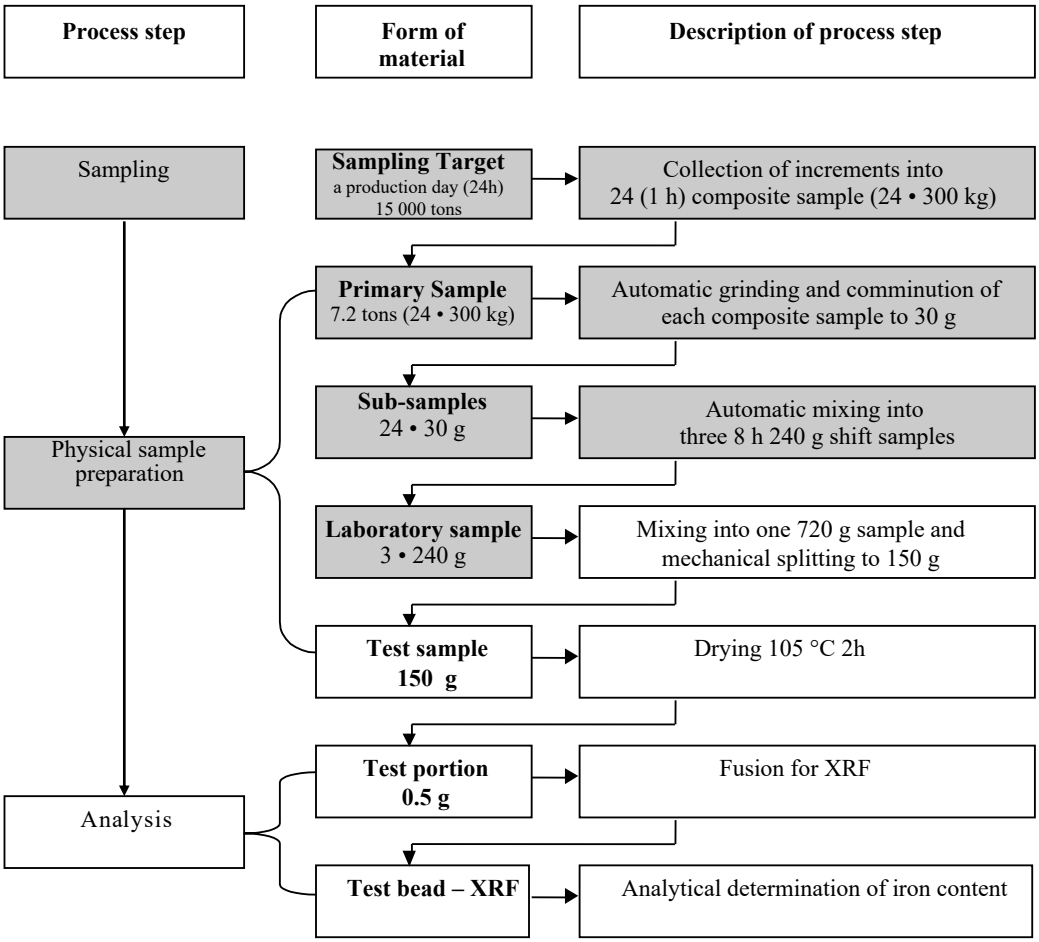


Figure B 1 Schematic diagram of iron ore sampling and analysis at LKAB, Kiruna

B4 Study design – double split

The study design was set up using ISO 3085 (1986) Method 1 (Figure B 2) as a template. The modified design used in this study is shown in Figure B 3. This design uses duplicate analyses of the three shift samples that represents the sampling target, i.e. the pellets produced in one production day. However the 3 shift samples are separated in time, so if the iron content varied significantly over a 24 hour period this would result in an overestimation of the sampling uncertainty. This issue is discussed in Section B7 below. This approach only takes into account precision. The overall analytical variation over time as well as any analytical bias is taken from the laboratory's analytical uncertainty estimation.

B4.1 Validation

The validation programme was set up using the protocol of ISO 3085 method 1 as a template shown in Figure B 2.

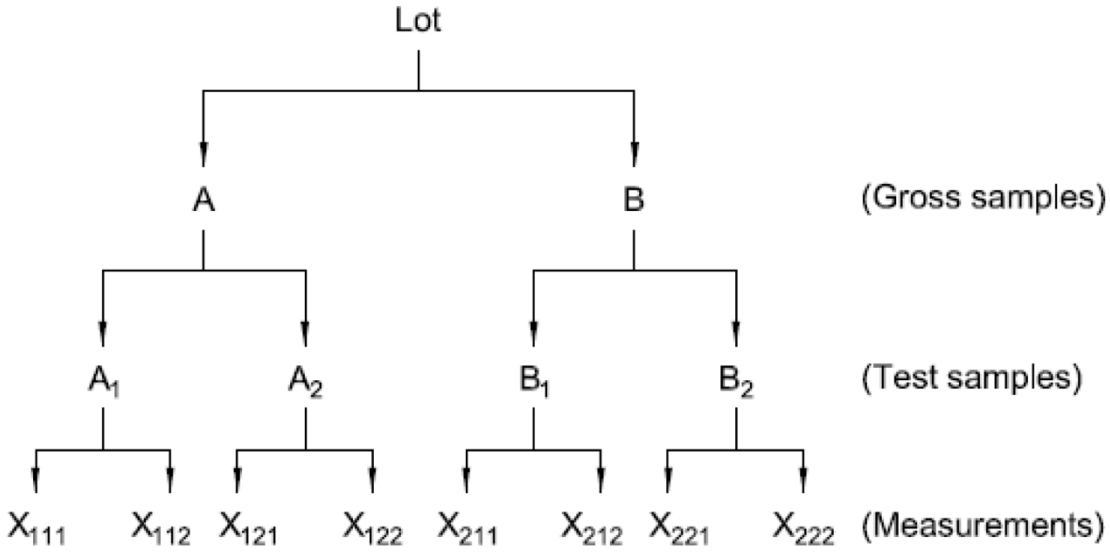


Figure B 2 Experimental design using duplicates – ISO 3085 (1986) Method 1

With separate sampling of the three shifts and no split of test sample and duplicate analyses the modified design used in this study is shown in Figure B 3.

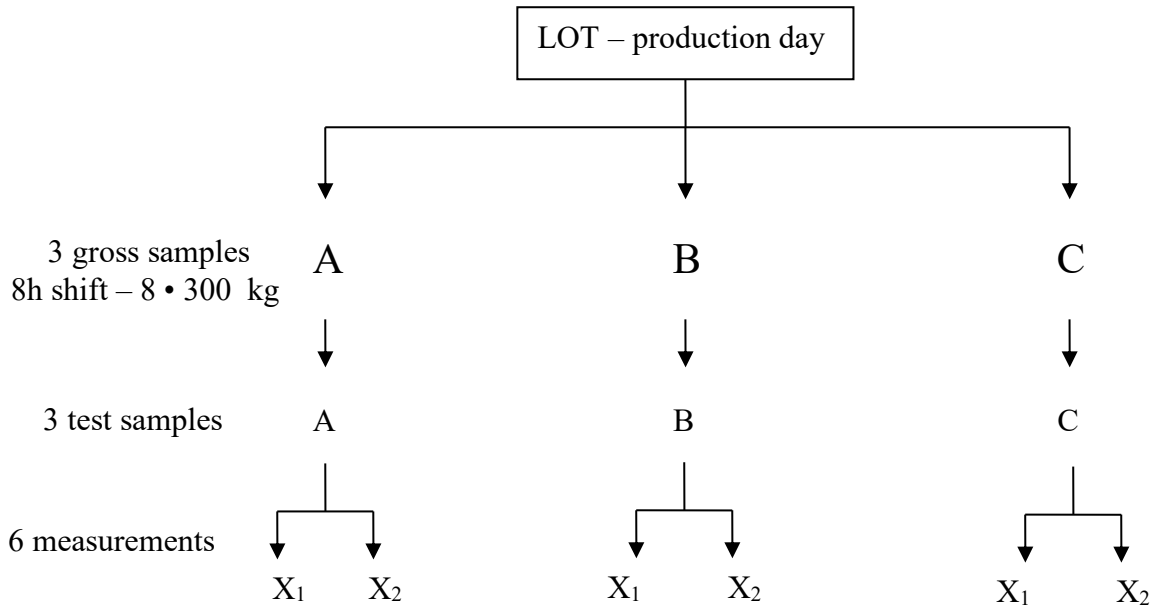


Figure B 3 Experimental design, modified method 1 of ISO 3085, separating the primary sample into three shift samples (240 g) and splitting to a test sample (150 g) , duplicate measurements of each test sample

B4.2 Quality control

The quality control programme can be set up by annual repetition of the validation experiment with three lot samples – a total of 18 analyses.

B5 Sample preparation and analysis

The laboratory sample (720 g) is split using a mechanical splitter and the test sample (150 g) is dried (105 °C, 2h). A test portion (0.5 g) is mixed with flux and fused into a glass bead, which is measured with XRF. The XRF instrument is calibrated with CRMs. The iron content is calculated by difference: 100 % minus impurities and minus oxygen. The expanded uncertainty reported by the laboratory is 0.20 % Fe at a level of 68 %.

B6 Results

B6.1 Range calculations and estimation of sampling standard deviation

The overall variation in production including analysis, sampling and product variation between September 2004 and November 2005) was 0.16 % Fe, expressed as one standard deviation at an iron level of 68 % Fe.

This standard deviation consists of the following parts

$$s_{total}^2 = s_{production}^2 + s_{samp}^2 + s_{anal}^2$$ and the measurement part is

$$s_{meas}^2 = s_{samp}^2 + s_{anal}^2$$

The results and calculations from data during one week in December 2005 are shown in Table B 1 to Table B 3. Raw data (*n = 42)* is given in Table B 4.

Table B 1 Range calculations for the analysis part – iron ore data Table B 4

Parameter	% Fe	Comment
Analysis – mean range from duplicates	0.046	
Analysis – stand dev. Estimated from range	0.041	*s* = range/1.128

The estimated analytical variation under repeatability conditions is *s* = 0.041 % Fe expressed as one standard deviation. The mean range of duplicate analyses is estimated to 0.046 % Fe. From duplicate measurements the standard deviation, 0.041 % Fe, is obtained by dividing the range with a factor of 1.128 when the range is based on duplicates (*n* = 2). This is then a standard deviation for a *single analytical* measurement.

Table B 2 Range calculations for measurement part – iron ore data Table B 4

Parameter	% Fe	Comment
Measurement – mean range from –triplicates	0.050	
Measurement – standard dev. from range	0.030	*s* = range/1.693
NOTE: Duplicate measurement on three separate 8 h shifts during a production day		

The estimated measurement variation under repeatability conditions is *s* = 0.030 % Fe. The mean range estimated is 0.050 % Fe. From triplicate measurements the standard deviation 0.030 % Fe is obtained by dividing the range with a factor of 1.693. This is then a standard deviation for a *single shift* measurement.

Table B 3 Calculations of the sampling part – iron ore data

Parameter	% Fe	Comment
Measurement – standard dev.	0.030	Measurement (sampling + analysis)
Analysis – standard dev.	0.041	Analytical part
Sampling – standard dev.	< 0.01	$s_{samp} = \sqrt{0.030^2 - \left(\frac{0.041}{\sqrt{3}}\right)^2}$

The sampling part of the variation, < 0.01 % Fe, is obtained using the following equation with three replicates – see Section 9.4:

$$s_{samp} = \sqrt{s_{meas}^2 - \left(\frac{s_{analysis}}{\sqrt{3}}\right)^2}$$

B6.2 Validation of analysis bias

Comparison with a long-time study from proficiency testing shows no significant analytical bias.

B6.3 Validation of sampling and analytical bias

Long-time studies comparing the results with customer measuring the iron content in the same lot show no significant bias (data obtained from LKAB, personal communication).

B6.4 Measurement uncertainty

The repeatability part of the expanded uncertainty (level of confidence of approximately 95 %) obtained in this study is 0.08 % Fe (2 · 0.041 % Fe). From the analytical laboratory at LKAB we obtain the within-lab reproducibility of the expanded uncertainty to be 0.14 % Fe. The expanded analytical uncertainty is estimated to be 0.20 % Fe. The expanded uncertainty for sampling < 0.02 % Fe (< 2 · 0.01 % Fe) and for measurement uncertainty 0.20 % Fe.

B7 Comments

For comparison, the calculations were also performed using ANOVA on four samples (with no missing data) between 14 November and 17 December 2005 with similar results. The s_{anal} from ANOVA is 0.039 % Fe and from range statistics 0.041 % Fe (n =1) and the sampling uncertainty is not significantly different from zero when using an F-test.

The estimated sampling uncertainty here is low. The drawback of estimations based on triplicates separated in time is that such a procedure could result in an overestimate due to production variations. In this case we can conclude that there is no overestimation, since the sampling uncertainty is estimated to be non-significant.

B8 Assessment of fitness for purpose of these measurements

With this low sampling uncertainty, this sampling procedure for determining Fe in iron ore pellets is fit for purpose. The low sampling uncertainty is obtained because the sampling equipment used here is designed for sampling of several parameters that are more heterogeneous, e.g. particle size distribution.

B9 Reporting and interpretation

An analytical result can be reported, e.g. Fe is 68.0 % ± 0.2 %.

B10 Summary

All values are expressed as uncertainty at a level of confidence of approximately 95 % of the iron concentration for the sampling target of one calendar day. The random part of the expanded analytical uncertainty is 0.08 % Fe and the random part of sampling uncertainty is <0.02 % Fe. In this case the random sampling uncertainty is less than half the random analytical uncertainty. However, the test is performed under one production week and sampling uncertainty may vary with production conditions.

The measurement uncertainty including sampling and analysis as well as random and systematic effects is estimated to be 0.20 % Fe.

Expanded uncertainty for a lot (24 h)			**Between target variability**
Sampling	**Analytical[1]**	**Measurement**	**Typical production variation[2]**
< 0.02 % Fe	0.20 % Fe	0.20 % Fe	0.16 % Fe
[1]Obtained from the analytical laboratory at LKAB. Estimated according to Nordtest TR 537 [33]. [2]Standard deviation measured under one year (September 2004 to November 2005).			

Acknowledgement
The authors are grateful for all assistance and data provided by LKAB, Kiruna, Sweden.

B11 Raw data for iron ore

Table B 4 Sampling one week in December 2005 according to modified method 1 of ISO 3085 – a lot split into three test samples and duplicate measurements of each sample

Lot date	Lot	Sample 1		Sample 2		Sample 3	
	x_i	x_{i11}	x_{i12}	x_{i21}	x_{i22}	x_{i31}	x_{i32}
	% Fe	% Fe	% Fe	% Fe	% Fe	% Fe	% Fe
2005-12-12	68.05			68.02	68.11	68.01	
2005-12-13	68.07	68.09		67.97	68.08	68.05	68.14
2005-12-14	68.11	68.03	68.15	68.09	68.11	68.16	68.14
2005-12-15	68.07	68.13	68.01	68.05	68.07	68.08	68.05
2005-12-16	68.06	68.05	68.08	68.09	68.04	68.05	68.06
2005-12-17	68.03	68.06	68.05	67.99	68.02	68.06	68.02
2005-12-18	68.02			68.03	68.00	68.03	68.02

Annex C Vitamin A in baby porridge

Measurand				Uncertainty estimation		
Analyte & technique	**Unit**	**Sector & matrix**	**Sampling target**	**Purpose**	**Design**	**Statistics**
Vitamin A (as retinol) HPLC	µg/100 g	Food baby porridge	Produced batch	Measurement	Balanced double split	ANOVA

C1 Scope

The scope is to estimate the measurement uncertainty and contributions from sampling and analyses. The estimates are based on samples from one type of baby porridge (see Table C 1) taken from 10 different batches, using a sampling procedure collecting duplicate samples from each batch.

C2 Scenario and sampling target

In the production of baby (infant) porridge, vitamin A (retinol) is added as a premix (together with vitamin D and vitamin C).[1] The premix is a minor ingredient. All ingredients are mixed thoroughly before distribution into packages. Earlier analysis indicated a bigger variation in analytical result between packages than expected. An expanded measurement uncertainty of 20 - 30 % would be considered acceptable. The question was raised if the variation mainly is due to analytical uncertainty or to sampling uncertainty. One of the theories suggests that the vitamin is locally unevenly distributed within the package, and therefore will give bigger analytical uncertainty if the test portion is too small[2] e.g. 3 - 5 g. A possible explanation of the heterogeneity is that the strongly lipophilic vitamin A molecules attach by electrostatic interactions to unevenly distributed lipophilic portions of the fruit particles in the porridge powder. The producers recommend a test portion size of 40 – 50 g whenever analysing vitamin A, D and C in baby porridge powder.

Table C 1 Product data provided by the producer Nestlé

Product data	**Oatmeal porridge with bananas and apricots**
Weight of batch, including premix (1 batch = 2 mixing containers)	1092 kg
Weight of added vitamin-premix in batch	1.228 kg
Vitamin A in premix (data from the Certificate of Analysis)	9016 IU/g = 2705 µg/g as retinol.
Vitamin A added to the batch	304 µg/100 g (retinol)
Vitamin A in ingredients according to the product specification	45 µg/100 g (retinol)
Estimated "true value" of Vitamin A[1]	349 µg/100 g (retinol)
NOTE: Vitamin A declared as Retinol – (Sum of trans- and cis-Retinol)	
[1]Data for estimating the "true value" of vitamin A in baby porridge are provided by the producer (Nestlé) of the product chosen for the validation.	

[1] Vitamin A is a group of unsaturated nutritional organic compounds that includes retinol, retinal, retinoic acid, and several provitamin. In this example retinol is determined.

[2]EN 12823-1 "Foodstuffs – determination of vitamin A by HPLC" indicates a test sample of approximately 2 to 20 g.

In order to compare the measured vitamin A concentration against declared values and European regulatory thresholds, an estimation of measurement uncertainty is desirable. To determine the random component of the measurement the double split design is chosen. To estimate the bias a comparison with the reference given in Table C 1 is made.

C3 Sampling procedure

Normally a spot sampling approach – one sample (one package) of a batch - is used as screening when comparing the content with declared values and legal limits.

Validation – In this study two samples are collected from each of 10 different batches of one type of baby porridge powder. Each sample is equal to one package of approximately 400 g powder.

Quality Control – Quality control (QC) of sampling from different types of baby porridge is done by collecting two samples from each of 8 batches of different types of baby porridges. All the types of porridges contain fruit in addition to milled cereals.

To ensure the quality in each package of the product at the time of the "best before date" of the porridge powder, the producer wraps the product in an airtight and light protecting bag. It is therefore assumed the degradation of the vitamin A is negligible during normal shelf life. The sampling for the validation was performed by the producer according to a specified procedure. For QC, the samples were purchased partly at the producers, partly at the retailer. When the samples were collected from retailers, care was taken to collect the two samples (of each product) at different retailers but in addition to assure the samples had the same batch marking. This is important to avoid adding batch variations to the apparent sampling distribution.

C4 Study design – double split replicates

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

C4.1 Validation

Samples are collected online (just after the filling operation of packages) at random time. Two samples (2 packages, each of approximately 400 g) are collected from each of ten production units (batches) of one type of baby porridge powder.

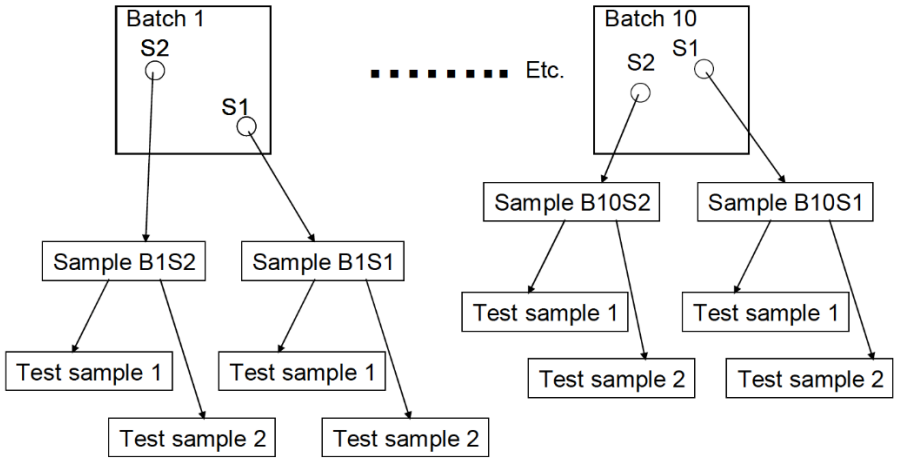


Figure C 1 Sampling for validation. Two samples are taken from each of ten production units/batches of the same type of sample

C4.2 Quality control

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

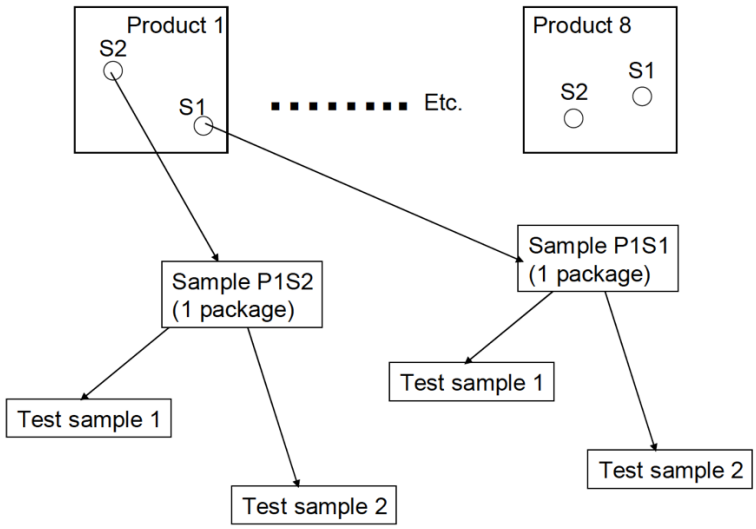


Figure C 2 Sampling for QC – two samples are taken from one batch of each of eight different types of baby porridge

C5 Sample preparation and analysis

The analytical work is done by "The National Institute of Nutrition and Seafood Research" (NIFES) according to method EN 12823-1. The laboratory is accredited according to ISO/IEC 17025.

The laboratory participates in proficiency testing schemes provided by FAPAS and BIPEA[1] with good results (in the period 2000 – 2005, $|z\ \text{score}| < 1$). The bias is

[1] PT provider in Great Britain and France respectively.

validated using a CRM. Data concerning the laboratory performance is given in Table C 2 below.

Table C 2 Method and performance data from quality control of vitamin A determined as retinol - laboratory analysis

Method	EN 12823-1 (HPLC – normal phase column - UV-detection)
Repeatability	CV = 3 %
Within-lab reproducibility	CV = 4 %
Expanded analytical uncertainty	14 %
Recovery	Standard addition, in laboratory: 90 – 110 % Based on laboratory PTs (in period 1999 – 2005), different matrixes: 88 – 113 %, mean recovery 100.5 %
Limit of Quantification (LOQ)	14 µg/100 g
CRM NIST 2383 – baby food 80 ± 15 µg/100 g (95 % confidence interval)	Laboratory result (*n* =28) 77 ± 14 µg/100 g (95 % confidence interval)

C5.1 Secondary sampling

A mechanical sample divider (Retsch) is used to split the primary samples – the packages . From each of the primary samples, 4 test samples are collected; two portions of approximately 3 - 5 g and two portions of approximately 40 – 50 g.

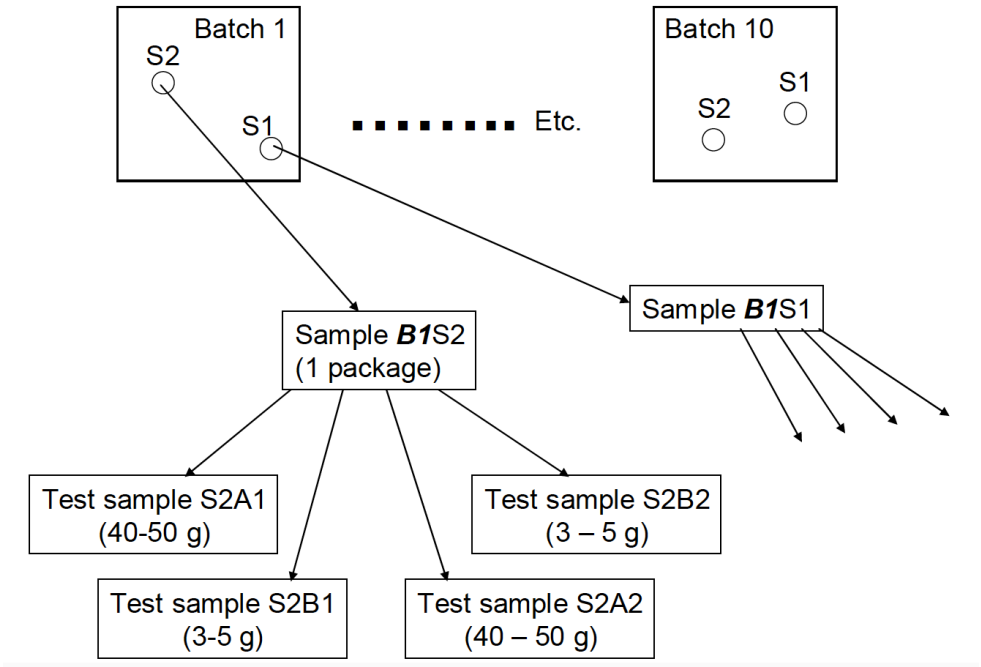


Figure C 3 Splitting of the primary sample to make 4 test samples

C5.2 Analysis

The analytical method is based on EN 12823-1 [36]. Retinol is saponified by using ethanolic potassium hydroxide containing antioxidants. Vitamin A is extracted by using hexane. Analysis is performed by using HPLC with UV detector.

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

C6 Results

C6.1 Test sample 40 g – baby porridge

Table C 3 Validation data – same product, results given in µg/100 g powder

Batch	S1A1	S1A2	S2A1	S2A2
B1	402	325	361	351
B2	382	319	349	362
B3	332	291	397	348
B4	280	278	358	321
B5	370	409	378	460
B6	344	318	381	392
B7	297	333	341	315
B8	336	320	292	306
B9	372	353	332	337
B10	407	361	322	382
NOTE1: S1 and S2: Primary samples from sampling location 1 and 2 of one production batch, A1 and A2. NOTE2: Mean value 348 µg/100 g, CV between batches is 6.1 %.				

C6.2 Test sample 4 g – baby porridge

Table C 4 Validation data – same product, results given in µg/100 g powder

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

NOTE1: S1 and S2: Primary samples from sampling location 1 and 2 of one production batch, B1 and B2: Analyses of duplicate test samples of a primary sample S.

NOTE2: Mean value 341 µg/100 g, CV between batches is 10.7 %.

C6.3 Calculations – test sample 40 g

In this study the calculations are done in an Excel spreadsheet and the details of the ANOVA calculations are shown in Box 11 and Box 12 in Section 9.5.

Calculation of uncertainty of analysis, ANOVA

Table C 5 Results from ANOVA calculations, 40 g test portion – analytical repeatability from sum of squares of differences, within groups (SS_{anal})

SS_{anal} $(\mu g/100g)^2$	(df_{anal})	s^2_{anal} $(\mu g/100g)^2$	s_{anal} $(\mu g/100g)$	CV_{anal} (%)
16 595	20	829.7	28.8	8.3
NOTE: For details see Box 11				

Calculation of sampling uncertainty, ANOVA

Table C 6 Results from ANOVA calculations, 40 g test portion – sampling repeatability from sum of squares of differences SS_{meas}

SS_{meas} $(\mu g/100g)^2$	(df_{meas})	s^2_{samp} $(\mu g/100g)^2$	s_{samp} $(\mu g/100g)$	CV_{samp} (%)
14 231	10	296.7	17.22	4.95
NOTE: For details see Box 12.				

Calculation of measurement uncertainty

The *CV* value from the ANOVA calculation can be used as an estimate of the standard uncertainty *u* (%). The analytical laboratory has estimated the analytical standard uncertainty to be 7 %, which is lower than the random analytical component for this

sample type, 8.28 %. The higher value of these two is used in the calculations. Combining the *CV* values from Table C 5 and Table C 6 with Equation 2, the results can be written as in Table C 7.

Table C 7 Measurement, sampling and analytical uncertainty – 40 g test sample

	Sampling %	Analytical %	Measurement %
Standard uncertainty u	4.95	8.28	9.7
Expanded uncertainty U	9.90	16.6	19

C6.4 Calculations – test sample 4 g

Calculation of analytical uncertainty, ANOVA

The same calculations are used as for test sample size of 40 g (see Box 11 and Box 12).

Table C 8 Results from ANOVA calculations, test portion 4 g – analytical repeatability from sum of squares of differences, within groups (SS_{anal})

SS_{anal} (µg/100g)^2	(df_{anal})	s^2_{anal} (µg/100g)	s_{anal} (µg/100g)	CV_{anal} (%)
312 206.5	20	15 610.325	124.9413	36.68

Calculation of sampling uncertainty, ANOVA

Table C 9 Results from ANOVA calculations, test portion 4 g – sampling repeatability from sum of squares of differences SS_{meas}

SS_{meas} (µg/100g)^2	df_{meas}	s^2_{samp} (µg/100g)	s_{samp} (µg/100g)	CV_{samp} (%)
102 860.25	10	-2662.15	Set to zero	-

The negative value of s^2_{samp} indicates that s_{samp} is small compared to the calculated value of s_{anal}. In this case, the estimates of s_{anal} and s_{samp} using robust ANOVA confirmed the smaller sampling standard deviation; the robust ANOVA estimates were: u_{samp} = 6.9 % and u_{anal} = 30 %. As the sampling is identical for the experiments with 40 g and 4 g test samples the sampling uncertainty should be the same, and a $CV_{samp} \approx 5$ % (see Table C 7) is used as an estimate.

Calculation of measurement uncertainty

Using the calculated *CV* value in Table C 8 and Table C 9 as an estimate of the measurement uncertainty and combining with Equation 2, the results can be written as follows:

Table C 10 Measurement, sampling and analytical uncertainty – 4 g test sample.

	Sampling (%)[1]	Analytical (%)	Measurement (%)
Standard uncertainty	4.95	36.7	37
Expanded uncertainty	9.90	73.4	74

[1]The u (%) value is derived from calculations using 40 g test samples

C6.5 Effect of the size of test sample on measurement uncertainty

The baby porridge powder looks homogeneous, and therefore a low measurement uncertainty is expected. However analyses of the powder indicated in fact a surprisingly large uncertainty when using a test sample size of 4 g and other commonly used methods often indicate a test sample size of approximately 2 – 20 g. The producers recommend using a test sample size of 40 – 50 g.

The validation tests gave the following results see Table C 11.

Table C 11 Comparing measurement uncertainty when analysing test samples of 40 g and 4 g

Test sample size	Expanded uncertainty U_{meas}
40 g test sample	19 %
4 g test sample	74 %

An U_{meas} of approximately 20 % is acceptable while an U_{meas} of 74 % is considered to be too high, taking into account the matrix and production conditions of this type of product. It can therefore be concluded that a test portion of 4 g is not "fit for purpose" when analysing vitamin A (retinol) in baby porridge powder containing milled cereals and fruit. A test portion of 40 – 50 g is recommended. This also supports the theory that the vitamin is unevenly distributed in the product, possible as local "hot spots" due to electrostatic interactions.

C6.6 Quality control

The quality control is here used for new batches of baby porridge to check if variation is similar in the new batches compared with estimated uncertainties. The construction of a range control chart is described in Section 8.28.2. In the case of baby porridge (40 g test sample) the following calculations can be made:

Action limit:	$AL = 3.69 \cdot \sqrt{4.95^2 + 8{,}28^2} = 36\ \%$
Warning limit	$WL = 2.83 \cdot \sqrt{4.95^2 + 8{,}28^2} = 27\ \%$
Central line	$CL = 1.128 \cdot \sqrt{4.95^2 + 8{,}28^2} = 11\ \%$

Table C 12 Quality control data (µg/100 g) with test portion 40 g – different products

Product	Producer	Porridge powder ingredients	S1A1	S1A2	S2A1	S2A2
P1	1	Oat, rice and pear	322	319	350	375
P2	1	Oat, rye, rice and pear	332	317	358	393
P3	1	Wheat, banana and apple	443	430	461	388
P4	1	Wheat and apple	318	383	390	334
P5	2	Oat, rice and banana	252	219	265	227
P6	2	Wheat and apple	274	239	233	217
P7	2	Oat, rice and apple	206	225	198	195
P8	3	Wheat, spelt, oat and apple (organic product)	392	335	375	416
NOTE: S1 and S2: Primary samples (laboratory samples) from sampling location 1 and 2 of one batch from each product. A1 and A2: Analyses on two test samples from each laboratory sample						

Table C 13 Quality control (µg/100 g): Calculation of differences D_{ik} and relative difference d_{ik} (%) between samples where (*i*) is the batch/product analysed, (*j*) is the number of samples from each batch and (*k*) is the test portion analysed

Product	Analyses	Sample S1 X_{i1k}	Sample S2 X_{i2k}	Difference D_{ik}	Mean	Rel. difference d_{ik}(%)
P1	A1	322	350	28	336	8
P2		332	358	26	345	8
P3		443	461	18	452	4
P4		318	390	72	354	20
P5		252	265	13	259	5
P6		274	233	41	254	16
P7		206	198	8	202	4
P8		392	375	17	384	4
P1	A2	319	375	56	347	16
P2		317	393	76	355	21
P3		430	388	42	409	10
P4		383	334	49	359	14
P5		219	227	8	223	4
P6		239	217	22	228	10
P7		225	195	30	210	14
P8		335	416	81	376	22

The relative difference *d* (%) can be compared directly with the action limit, or is presented in a control chart, see Figure C 4.

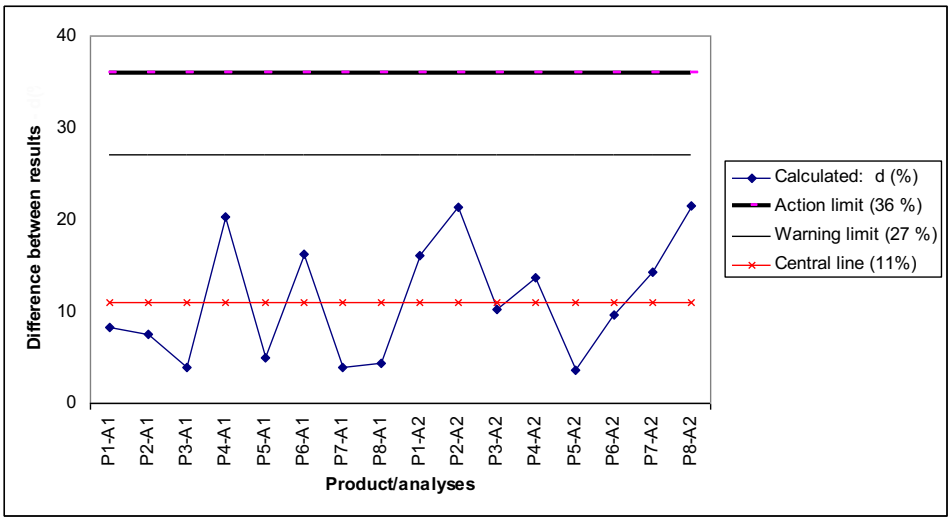


Figure C 4 Control chart, QC analyses of vitamin A in baby porridge containing cereals and fruits

The control chart in Figure C 4 shows that when collecting duplicated samples from the same batch, the difference between analytical results d (%) is smaller than the action limit AL. All the calculated differences are in fact smaller than the calculated warning limit, WL= 27 %.

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

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

C6.7 Measurement uncertainty

Sampling uncertainty

Calculations from the validation study gave for 40 g test sample an expanded sampling uncertainty, U_{samp} = 9.9 % (see Table C 7). The calculated uncertainty includes only repeatability.

Analytical uncertainty

Calculation from the validation study gave an expanded analytical uncertainty, U_{anal} = 17 % for the 40 g test sample. The laboratory reports their own estimation of the analytical uncertainty (see Table C 2): $2 \cdot CV_{inlab}$ = 14 %. The value of $2 \cdot CV_{inlab}$ is used as an estimate of U_{anal} in the laboratory. The U_{anal} found in the validation study was at the same level but still a little bigger than the U_{anal} reported by the laboratory.

Calculations from the validation study gave a relative expanded uncertainty U_{meas} = 19 % ~ 20 % (40 g test sample – see Table C 7).

Bias

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

Table C 14 Certified and analysed data (retinol) for CRM 2383

CRM 2383	Mean value (µg/100 g)	*U* (µg/100 g)	Bias (%)
Certified	80	15	-
Analysed	77	14	- 3.75

The measurement uncertainty and the bias determined for the CRM could be allowed for in the analytical measurement uncertainty as in example A4 of the Eurachem Guide [24]. The laboratory normally reports a recovery of 90 – 110 %. Recovery based on laboratory PTs 1999-2005 is the range 88 – 113 %. The results for the PT indicate no, or a very small, bias. Analyses of CRM 2383 in the laboratory give a mean value of 96.3 % of the certified value – which indicates a small bias (-3.7 %). As the matrix of the CRM "baby food composite" is different to the baby porridge, and the analytical method includes an extraction, the bias determined when analysing the CRM might not be representative for the analyses of baby porridge. In this Handbook we decided to always try to use the analytical uncertainty given by the laboratory that includes within-lab reproducibility and uncertainty of the bias – in this case 14 %.

In the validation study, the mean value of retinol was determined to be 348 µg/100 g (when using a test sample of 40 g). According to data provided by the producer (see Table C 1), the "true value" for retinol was calculated to be 349 µg/100 g porridge powder. This gives a recovery > 99 % of the "true value". *This gives an indication that the systematic error due to sampling and analyses is small and might be negligible when analysing baby porridge-powder containing milled cereals and fruits – on the condition that a test sample of at least 40 – 50 g is used.*

C7 Comments

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

When baby porridge-powder containing milled cereals and fruit is to be analysed, it is recommended to use a relatively large test sample of approximately 40 – 50 g and not 2 – 20 g as often indicated in test methods.

C8 Assessment of fitness for purpose

The measurement uncertainty is acceptable and therefore the sampling procedure is fit for purpose. However, a test sample size of at least 40 - 50 g should be used, otherwise the analytical method used is not fit for purpose.

C9 Reporting and interpretation

The analytical result of retinol in baby porridge should be reported as the determined value *x* with the expanded uncertainty: *x* ± 20 %.

C10 Summary

Expanded Uncertainty			Between-target variability *CV*
Sampling	Analytical	Measurement	
9.9 %	16.6 %	19 %	6.1 %
NOTE: Calculated with data given in Table C 3.			

Acknowledgement

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

Annex D Electrical conductivity in industrial wastewater

Measurand				Uncertainty estimation		
Analyte & technique	**Unit**[1]	**Sector & matrix**	**Sampling target**	**Purpose**	**Design**	**Statistics**
Electrical conductivity	mS/m	Industrial wastewater	Wastewater outlet spot samples	Sampling and analysis	Time series	Variographic analysis
[1]The unit is milliSiemens per metre						

D1 Scope

In this example the data will be evaluated using so-called variographic analysis as described in Section 9.7. The scope is to estimate the measurement uncertainty as well as individual uncertainty contributions from inherent heterogeneity, the automatic sampling, pretreatment and analyses of wastewater in relation to the measurement of electrical conductivity. The intension is to focus on the uncertainty contribution from sampling using specific wastewater sampling equipment.

D2 Scenario and sampling target

Sampling and analysis of wastewater may be carried out for a number of reasons, typically:

- for investigations related to specific control limits (industrial wastewater);
- for monitoring of inlet to wastewater treatment plants for optimisation of the wastewater treatment process;
- for surveillance of the outlet from an industry or wastewater treatment plant related to allowable limits;
- for supervision of the treatment processes.

Quality characterization of a wastewater stream aims to determining the concentration or load of pollutants in the wastewater, generally during an extended period of time, for example 1) to monitor compliance with a control limit, 2) to determine trends, 3) to provide data on unit process efficiency or 4) to provide loading data for planning and/or design purposes.

Fees and fines on wastewater pollutant loads are often based on the results from sampling and analyses of the specific water streams. Failure to conduct proper sampling and analyses may result in problems in the management of the wastewater treatment plant and/or severe environmental problems, as well as it may result in non-justified economical burdens for the wastewater producer due to incorrect fees being generated. It is, therefore, in the interest of both the wastewater producer and the supervising authority to assure a uniform and representative sampling (i.e. sampling bias is assumed to be negligible) and uniform and reproducible results with a known and acceptable uncertainty.

D3 Sampling procedure

Sampling of wastewater is conducted using the procedures as described in the standard ISO 5667-10 Water quality – Sampling Part 10: Guidance on sampling of wastewaters. The standard gives guidance on the selection of the sampling point to assure representative sampling. It describes manual sampling as well as automatic sampling of wastewater. The present study represents sampling using automatic equipment, where the principle is that the sampler takes a series of discrete samples at fixed intervals and held in individual containers. In practical cases the same design is used when carrying out 24-hour studies to identify peak loads.

The same equipment can be adjusted to take flow-proportional samples, where the frequency or volume of sampling is adjusted according to the variations in the flow of wastewater, each held in individual containers and can furthermore be adjusted to take time-dependent and flow-proportional composite samples. The most frequent practice for control of wastewater is the use of the principle where flow-proportional composite samples are taken over 24 hours.

D4 Study design

In this example results for electrical conductivity in the samples from an industrial wastewater outlet are presented. Wastewater was sampled by the use of automatic wastewater sampling equipment. Spot samples were taken at equal time-intervals during preselected periods. Each spot sample was analysed for electrical conductivity.

To find out the uncertainty arising from the sampling process we consider the following equation, which says that the random part of total measurement uncertainty, here called s_{meas}, is the sum of the sampling uncertainty and the analytical uncertainty added by the variances, see Equation 2:

$$s_{meas}^2 = s_{samp}^2 + s_{anal}^2$$

Thus, if we can estimate the measurement uncertainty (s_{meas}) and the analytical uncertainty (s_{anal}), we will be able to estimate by calculation the part of the uncertainty that arises from the sampling process. The estimation of uncertainty from pretreatment and analyses of the wastewater samples (s_{anal}) is based on multiple treatment and analyses of samples taken at sites, but can also be estimated from data from internal quality control of laboratory analyses.

The time-series were analysed using the variographic analysis technique. For a more detailed description of the variographic analysis technique, see Section 9.7. Two series of increments were taken at each of the selected sampling points by using the same automatic sampling equipment. One of the series was repeated at each point:

1) A first series (denoted W) of 24 increments taken at constant interval (one hour) over 24 hours to study the variations in inorganic constituent represented by the electric conductivity;
2) A second series (carried out in duplicate, denoted X and Y) of 24 increments taken at constant interval (2.5 minutes) over 60 minutes. The individual spot samples were taken as closely together as possible with the given wastewater sampling equipment. The purpose of this series was to calculate an accurate estimate of the ordinate *V(0)* representing s_{meas} at the origin.

For Quality Control (QC), i.e. to calculate the s_{anal}, a 10 L sample of wastewater was sampled from the wastewater stream at the end of the sampling periods. The sample bottle was shaken, and the water distributed in 10 bottles for electric conductivity measurement. In addition all measurements were subjected to ordinary internal quality control by parallel analyses of synthetic quality control samples.

D5 Sample preparation and analysis

The sampling and measurements were carried out by Eurofins Environment A/S, which is accredited according to ISO/IEC 17025 for sampling of wastewater and laboratory analyses of the conductivity.

D5.1 Automatic sampling of wastewater

All samples have been taken using a fractionated time proportional sampling. The volumes of the discrete samples taken were 3×170 ml collected into on single 500 ml sample. It was deemed necessary to have a sample volume of 500 ml due to latter splitting of the sample for analysis of different chemical parameters.

For the sampling portable equipment from EPIC was used. This equipment is based on the vacuum principle and makes it possible to take up to 24 fractioned samples. Before and after each 24-hour period the equipment was used to take 24 samples over a 60-minute period, without making any changes to the installation (such as suction height, volume, sampling location etc.). It was deemed necessary to use 2.5 minutes intervals between the samples, allowing enough time for flushing the lines and the sample container. A period of 2.5 minutes between each sample was very near the absolute minimum time for the particular equipment.

The samples were taken and stored in the sampling equipment at ambient temperature. Immediately after each 24-hour period the samples were transported to the laboratory. The transportation time was about 45 minutes.

D5.2 Analysis

The analyses were carried out at Eurofins' laboratory in Vallensbæk. The analytical method used for conductivity was DS 288 (probe method).

D6 Results

The resulting data was collected and plotted in time series and as variograms in Figure D 1 to Figure D 6. Using the variograms for the 2.5 minutes time series X and Y, it was possible to estimate the *V(0)* or smallest possible measurement uncertainty (corresponding to the standard deviation, $s_{meas,}$ according to Equation 31), which in this case would include the uncertainty from the inherent heterogeneity of the samples, the sampling process and sample handling and analysis. The data, calculations and results are shown in Table D 2 for conductivity in the wastewater outlet from an industrial plant.

The results for series W, the 24-hour experiment, are shown as a time series in Figure D 1 and as a variogram in Figure D 2. For detailed information on how to construct a variogram, see Section 9.7 in the main text of this Handbook.

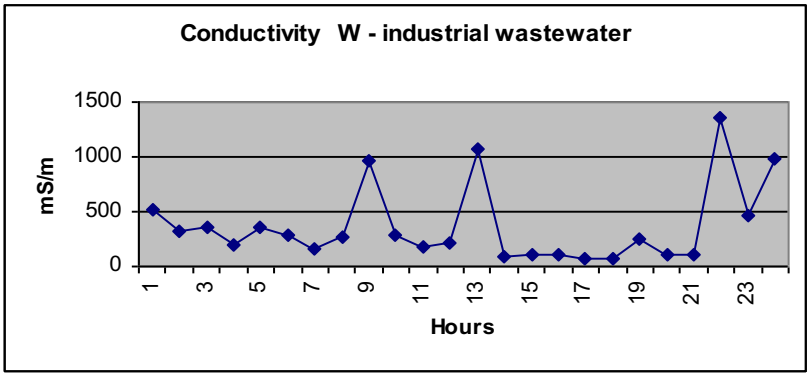


Figure D 1 Time series W (1 hour increments) – Conductivity

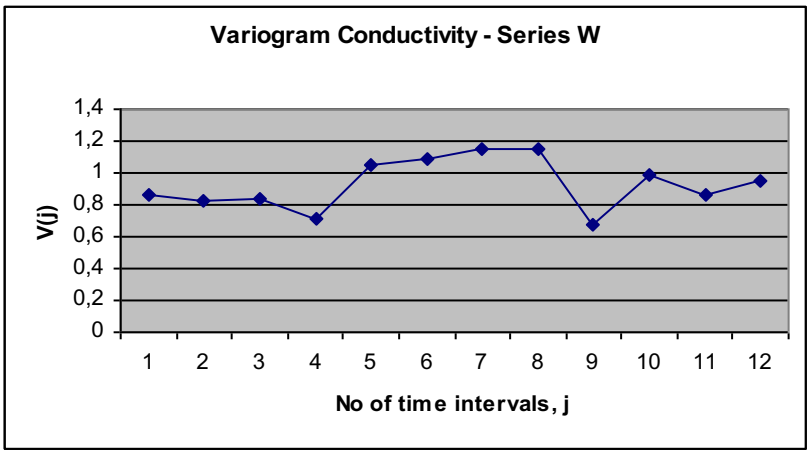


Figure D 2 Variogram of time series W (1 hour increments)

The 24-hour time series and the corresponding variogram reveal no hidden or unexpected structures in the conductivity concentrations over the 24-hour period, even though there is a small indication of a periodic cycle of 4-5 hours in the first part of the variogram. No conclusions on this should be drawn from a single experiment, but it is something that might be interesting to investigate in the future.

Figure D 3 and Figure D 4 show the results from the first of the two experiments with 2.5-minute intervals over 60 minutes, series X, sampled just before the 24-hour experiment in series W. The corresponding results for series Y, sampled just after series W, are shown in Figure D 5 and Figure D 6. The variograms of series X and Y showing the first 12 points only, makes it possible to estimate the smallest possible sampling error, $V(0)$ representing s_{meas}, from a fitted straight line through the points. This represents the smallest error, which would result if two samples could be taken with an infinitely small time distance between them. Note that the fitted straight lines in these variograms are constructed using the first 12 points only, since the uncertainty rises due to decreasing degrees of freedom for the latter points (for j=23 one single point determines the $V(23)$).

The relatively small variation in conductivity during the first part of series Y is also reflected in the variogram, and the estimate of $V(0)$ hence becomes smaller than in series X. Since no sampling variables were changed between the series, this almost certainly reflects variations in composition of the sample stream due to increased heterogeneity or short-term production variability.

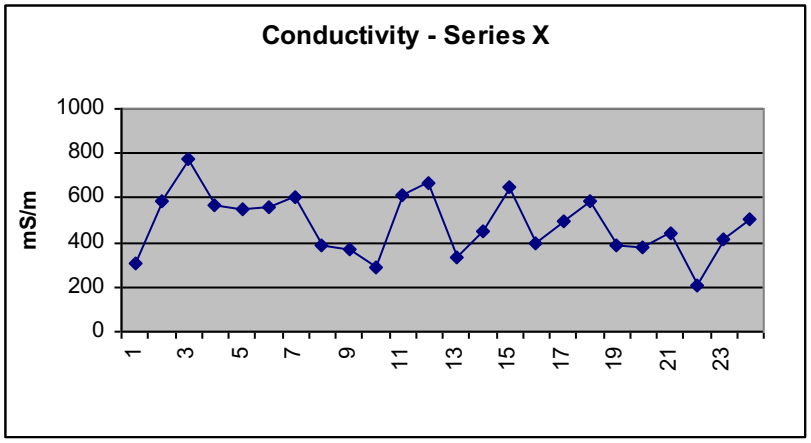


Figure D 3 Time series X (2.5 minutes increments)

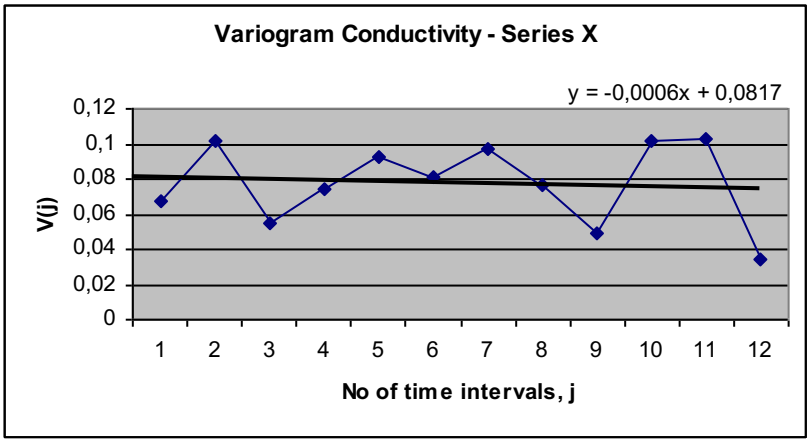


Figure D 4 Variogram of time series X (2.5 minutes increments)

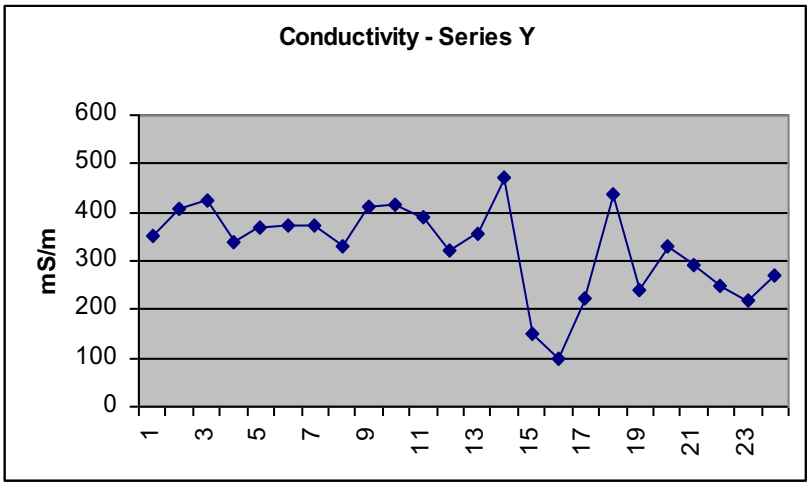


Figure D 5 Time series Y (2.5 minutes increments)

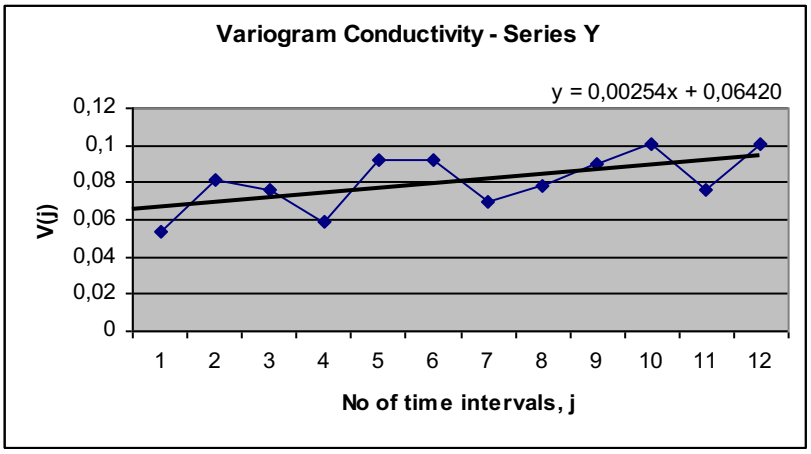


Figure D 6 Variogram of time series Y (2.5 minutes increments)

Using the information obtained from the fitted lines in the variograms, *V(0)* can be recalculated to the estimate *s(0)* (or s_{meas}) according to Equation 31 and CV according to Equation 32. The results of the three experiments and the calculations are summarised in Table D 1.

Table D 1 Summary of the experimental results of measurement of electrical conductivity from industrial wastewater outlet

	Mean mS/m	***s* mS/m**	**CV %**	***V(0)***	***s(0)* = s_{meas} mS/m**	**CV_{meas} %**
24 hours (W)	371	356	96		-	-
60 minutes (X)	481	138	29	0.0817	138	29
60 minutes (Y)	326	93	29	0.0642	83	25
Repeated analysis on one sample	340	4.4	1.3		-	-
Systematic bias QC		Negligible				

As we know the measurement uncertainty and the analytical uncertainty from our experiments, we now have enough information to be able to calculate the uncertainty arising from the sampling. The sampling uncertainty is calculated according to :

$$s_{samp} = \sqrt{s_{meas}^2 - s_{anal}^2}$$

V(0) is a representation of s_{meas}, and is taken directly from the linear regression of the variogram for 2.5-minute intervals. From the table above we can see that there are 2 different results for s_{meas}, from experiments X and Y. In the calculations the higher number from series X (s_{meas} = 138 mS/m) is used, in order not to underestimate the uncertainty. If we add information about the analytical uncertainty, in this case a *CV* of 1.3 % taken from repeated analysis of the 10 L samples, we can estimate the uncertainty from the sampling: $s_{anal} = 0.013 \cdot 481 = 6.07$ mS/m and $s_{samp} = \sqrt{138^2 - 6.07^2} = 138$ mS/m, corresponding to 29 % of the mean 481 mS/m.

Sampling thus contributes to virtually all the measurement uncertainty of each of the spot samples, and the analytical uncertainty is thus insignificant. In the present case the sample was taken at an industrial wastewater outlet with significant amounts of organic matter and particles in the sampling well. The sampling site did not have an optimal design, for example the lift height was rather high. The efficiency of mixing as well as the design of the sampling site are expected to be reflected in the uncertainty of measurement in spot samples.

It should also be kept in mind that certain additional uncertainty components are not treated in a single investigation like this, e.g. the uncertainty arising from repeated set-up of the sampling equipment. Furthermore, sampling bias is not included (i.e. does the sample taken truly represent the average concentration of the whole wastewater stream in the period under investigation?).

D7 Comments

The results clearly indicate that it is relevant to perform experiments to evaluate the uncertainty contributions from sampling, and not only the analysis, and that the uncertainty originating in the sampling step has to be evaluated individually for each location and sampling set-up. In sampling sites where the particle load is smaller and the mixing is better, the analytical error might, and has been shown, to correspond more significantly to the total uncertainty of the spot sample.

D8 Assessment of fitness for purpose

The sampling uncertainty is high, 29 % of the average concentration, for electric conductivity in the industrial wastewater from the site investigated. Sampling uncertainty dominates that total uncertainty and the analytical uncertainty is negligible in this context. The high uncertainty probably reflects that the design of the sampling site is not optimal and that the sampling target in heterogeneous. The study can therefore be used to identify and quantify the effect of sub-optimal design of the sampling site.

D9 Reporting and interpretation

Analytical results from the present site give the level of concentration but results should be interpreted with caution due to the high sampling uncertainty. Results could be reported as the determined value x with a relative expanded measurement uncertainty, coverage factor 2: x ± 60 %.

D10 Summary

The measurements in the outgoing industrial wastewater show that the contribution to the measurement uncertainty in the spot samples from the inherent heterogeneity and the sampling is the totally dominating source of uncertainty, and that the uncertainty from the analytical steps is insignificant in comparison. The main reason appears to be that the water is not well mixed in the sampling well or that the sampling site in some way is sub-optimal in the design, but the variability of the particle load might also be an important source of measurement uncertainty.

It should be noted that the results are valid for spot samples only and with the current sampling equipment and the current design of the sampling experiment only, and that factors not investigated might have further influence on the uncertainty.

The total relative standard deviation (calculated as the sum of the sampling variability and the analytical variability) for measurements of conductivity in a spot sample of the wastewater is estimated to 29 %.

Expanded uncertainty			**Target variability**
Sampling	**Analytical**	**Measurement**[1]	
58 %	2.6 %	60 %	-
[1]Expanded uncertainty 58 % rounded to 60 %			

D11 Raw data

Table D 2 Results of measurement of electric conductivity (EC) in water sampled by an automatic water sampler – series W sampling of over 24 h and series X and Y sampling of over 60 minutes

Time (hours)	**EC - Serie W** (mS/m)		**Time** (minutes)	**EC - Serie X** (mS/m)	**EC - Serie Y** (mS/m)
1	512		2.5	307	350
2	318		5	585	406
3	353		7.5	771	424
4	197		10	572	339
5	364		12.5	553	369
6	284		15	560	371
7	165		17.5	605	373
8	273		20	386	328
9	963		22.5	370	412
10	292		25	291	416
11	170		27.5	614	391
12	210		30	670	322
13	1069		32.5	329	354
14	89.7		35	452	470
15	108.5		37.5	649	148
16	110.4		40	397	99.8
17	77.4		42.5	494	221
18	80.3		45	583	437
19	251		47.5	386	239
20	100		50	381	332
21	105.1		52.5	443	292
22	1350		55	210	247
23	464		57.5	416	218
24	991		60	509	268
Mean	371		Mean	481	326
s	356		*s*	138	93
CV	96 %		CV	29 %	29 %

NOTE: This page is intentionally left blank.

Annex E Terminology

Analyte	Substance or parameter subject to measurement.
Bias	Estimate of a systematic measurement error VIM [16]
Composite sample (also average and aggregate)	Two or more increments/sub-samples mixed together in appropriate portions, either discretely or continuously (blended composite sample), from which the average value of a desired characteristic may be obtained. AMC [26]
Duplicate (Replicate) sample	One of the two (or more*) samples or sub-samples obtained separately at the same time by the same sampling procedure or sub-sampling procedure *for replicate sample *Note: each duplicate sample is obtained from a separate 'sampling point' within the 'sampling location'* AMC [26]
Homogeneity	The degree to which a property or constituent is uniformly distributed throughout a quantity of material. *Note 1. A material may be homogenous with respect to one analyte or property but heterogeneous with respect to another* *Note 2. The degree of heterogeneity (the opposite of homogeneity) is the determining factor of sampling error* IUPAC [28]
Increment	Individual portion of material collected by a single operation of a sampling device IUPAC [28], AMC [26]
Laboratory sample	Sample as prepared for sending to the laboratory and intended for inspection or testing. ISO 78-2 [29]
Measurand	Quantity intended to be measured VIM [16] *Authors' note: The specification of measurand regarding sampling target, analyte, unit and base for reporting is discussed in Section 4.2.*

Precision

Closeness of agreement between indications or measured quantity values obtained by replicate measurements on the same or similar objects under specified conditions

NOTE 1 Measurement precision is usually expressed numerically by measures of imprecision, such as standard deviation, variance, or coefficient of variation under the specified conditions of measurement.

NOTE 2 The 'specified conditions' can be, for example, repeatability conditions of measurement, intermediate precision conditions of measurement, or reproducibility conditions of measurement

VIM [16]

Authors' note: The term "within-laboratory reproducibility" is used in this Handbook to describe "intermediate precision conditions" since this term is used in other Nordtest handbooks. .

Primary sample

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

IUPAC [28], AMC [26]

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

Random sample

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

ISO 3534-1 [25]

Random sampling; simple random sampling

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

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

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

ISO 7002 [31]

Reference sampling

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

IUPAC [32]

Reference sampling target (RST)	The analogue in sampling of a reference material or certified reference material (in chemical analysis) *Note: A sampling target, one or more of whose element concentrations are well characterized in terms of spatial/time variability. The analogue in sampling of a reference material or a certified reference material (in chemical analysis) (note adapted from IUPAC (2003) draft recommendations; originally defined in ISO Guide 30: 1992)* Thompson and Ramsey [27]
Representative sample	Sample resulting from a sampling plan that can be expected to reflect adequately the properties of interest in the parent population IUPAC [28], AMC [26]
Sample	A portion of material selected from a larger quantity of material IUPAC [28], AMC [26]
Sample preparation	The set of material operations (such as reduction of sizes, mixing, dividing, etc.) that may be necessary to transform an aggregated or bulk sample into a laboratory or test sample *Note: The sample preparation should not, as far as possible, modify the ability of the sample to represent the population from which it was taken* Adapted from ISO 3534-1 [25]
Sample pretreatment	Collective noun for all procedures used for conditioning a sample to a defined state which allows subsequent examination or analysis or long-term storage Adapted from ISO 11074-2 [15]
Sample size	Number of items or the quantity of material constituting a sample ISO 11074-2 [15]
Sampler	Person (or group of persons) carrying out the sampling procedures at the sampling point *Note: The term 'sampler' does not refer to the instrument used for sampling, i.e. the 'sampling device'* Adapted from ISO 11074-2 [15]
Sampling	Process of drawing or constituting a sample *Note: For the purpose of soil investigation 'sampling' also relates to the selection of locations for the purpose of in situ testing carried out in the field without removal of material (from ISO 1998)* ISO 11074-2 [15]

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

AMC [26]

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

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

AMC [26]

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

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

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

AMC [26]

Authors' note: In this guide the term sampling repeatability, s_{samp} is used.

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

AMC [26]

Authors' note: In this guide sampling procedure is the detailed instruction for sampling in line with the VIM [16] definition of a ***measurement procedure****: detailed description of a measurement according to one or more measurement principles and to a given measurement method, based on a measurement model and including any calculation to obtain a measurement result.*

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

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

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

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

AMC [26]

Sub-sample	A sample taken from a sample of a population *Note 1. It may be selected by the same method as was used in selecting the original sample, but need not be so,* *Note 2. In sampling from bulk materials, sub-samples are often prepared by sample division. The sub-sample thus obtained is also called a "divided sample"* ISO 3534-1 [25]
Sub-sampling (Sample division)	Process of selection one or more sub-samples from a sample of a population ISO 11074-2 [15]
Test portion	Quantity of material, of proper size for measurement of the concentration or other property of interest, removed from the test sample IUPAC [28]; AMC [26]
Test sample	Sample, prepared from the laboratory sample, from which the test portions are removed for testing or analysis IUPAC [28]; AMC [26]
Uncertainty (of measurement)	Parameter, associated with the result of a measurement, that characterises the dispersion of the values that could reasonably be attributed to the measurand *Notes 1. The parameter may be, for example, a standard deviation (or a given multiple of it), or the half width of an interval having a stated level of confidence* *Note 2. Uncertainty of measurement comprises, in general, many components. Some of these components may be evaluated from the statistical distribution of the results of series of measurements and can be characterised by experimental standard deviations. The other components, which can also be characterised by standard deviations, are evaluated from assumed probability distributions based on experience or other information* *Note 3. It is understood that the result of the measurement is the best estimate of the value of the measurand, and that all components of uncertainty, including those arising from systematic effects, such as components associated with corrections and reference standards, contribute dispersion* JCGM 100 [30] *Authors' note: If measurand is defined in terms of the quantity within the sampling target, then uncertainty from sampling is included within uncertainty of measurement.*

Uncertainty factor	The factor by which the measured value is multiplied and divided in order to generate the limits of an uncertainty interval [38]
Uncertainty from sampling	The part of the total measurement uncertainty attributable to sampling IUPAC [32]

Bibliography

1. Commission directive 2009/90/EC, laying down, pursuant to Directive 2000/60/EC of the European Parliament and of the Council, technical specifications for chemical analysis and monitoring of water status https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:201:0036:0038:EN:PDF.
2. W. Horwitz (1980) Quality Assurance in the Analysis of Foods for Trace Constituents, Journal of the AOAC 63:6, 1344-1354.
3. M. Thompson (2000) Recent trends in inter-laboratory precision at ppb and sub-ppb concentrations in relation to fitness for purpose criteria in proficiency testing, Analyst 125, 385-386.
4. Commission Directive 2004/16/EC: Sampling methods and the methods of analysis for the official control of the levels of tin in canned foods.
5. Bekendtgørelse om kvalitetskrav til miljømålinger udført af akkrediterede laboratorier, certificerede personer m.v.(2006), https://www.retsinformation.dk/eli/lta/2006/1353.
6. Water Research Centre (1989), Manual on Analytical Quality Control for the Water Industry, TR66, ISBN 0902156853.
7. K. Esbensen (2006), Aalborg University Esbjerg, proposal for sampling method.
8. P.L. Smith (2001) A primer for sampling solids, liquids, and gases. ASA, SIAM.
9. S. Barbizzi, P. de Zorzi, M. Belli, A. Pati, U. Sansone, L. Stellato, M. Barbina, A. Deluisa, S. Menegon and V. Coletti (2004), Characterisation of a reference site for quantifying uncertainties related to soil sampling, Environmental Pollution 127 121-135.
10. EN ISO 8130-9 (1999), Coating powders – Part 9: Sampling, CEN.
11. EN ISO/IEC 17025 (2017), General requirements for the compentence of testing and calibration laboratorier, CEN.
12. EN ISO/IEC 17024 (2012), Conformity assessment – Generel requirements for bodies operating certification of persons, CEN.
13. M.H. Ramsey, (1993), Sampling and analytical quality control (SAX) for improved error estimation in the measurement of heavy metals in the environment, using robust analysis of variances, Applied Geochemistry 2, 149-153.
14. B. Magnusson, H. Hovind, M. Krysell, U. Lund and I. Mäkinen, Handbook - Inter Quality control, Nordtest Report TR 569 (5th ed.) 2018. Available from www.nordtest.info.
15. ISO 11074-2 (2015), Soil quality – Vocabulary – Part 2: Terms and definitions related to sampling, ISO, Geneva.
16. JCGM 200:2012 International vocabulary of metrology – Basic and general concepts and associated terms (VIM) 3rd edition. Available from www.bipm.org.
17. NMKL procedure no. 12 (2002), Guide on sampling for analyses of food.
18. M. Thompson, and R.J. Howarth, (1976), Duplicate Analysis in Geochemical Practice, Analyst 101: 690-698.
19. A.J. Duncan, (1974), Quality Control in Industrial Statistics, Richard D. Irwin Inc.

20. J.N. Miller and J.C. Miller (2010), Statistics and Chemometrics for Analytical Chemistry (6th ed.), Prentice Hall., ISBN 978-0-27-373042-2.
21. RANOVA – An AMC software for ANOVA and robust ANOVA calculations. Freely available from: https://www.rsc.org/Membership/Networking/InterestGroups/Analytical/AMC/Software/RANOVA.asp.
22. M.H. Ramsey (1998), Sampling as a source of measurement uncertainty: Techniques for quantification and comparison with analytical sources, Journal of Analytical Atomic Spectrometry 13, 97-104.
23. Miljøstyrelsens Referencelaboratorium (2005), Usikkerhed/fejl ved automatisk prøvetagning af spildevand. Litteraturundersøgelse og forsøgsskitse, Rapport til Miljøstyrelsen.
24. M.H. Ramsey and S.L.R. Ellison (eds.) Eurachem/EUROLAB/CITAC /Nordtest/AMC Guide: Measurement uncertainty arising from sampling: a guide to methods and approaches. 2nd Edition, Eurachem (2019). Available from www.eurachem.org.
25. ISO 3534-1 (2005), Statistics – Vocabulary and symbols, ISO, Geneva**.**
26. AMC (2005), Analytical Methods Committee Technical Brief No. 19. Terminology – the key to understanding analytical science. Part 2: Sampling and sample preparation. www.rsc.org/Membership/Networking/InterestGroups/Analytical/AMC/TechnicalBriefs.asp.
27. M. Thompson and M.H. Ramsey (1995), Quality concepts and practices applied to sampling – an exploratory study, Analyst 120, 261 – 27.
28. IUPAC (1990), Nomenclature for sampling in analytical chemistry. Prepared for publication by Horwitz, W., Pure and Applied Chemistry 62, 1193 – 1208.
29. ISO 78-2 (1999), Chemistry – Layout for standards – Part 2: Methods of chemical analysis. ISO, Geneva.
30. JCGM 100:2008 Evaluation of measurement data - Guide to the Expression of Uncertainty in Measurement (GUM). Sevres, (2008). www.bipm.org/en/publications/guides/gum.html.
31. ISO 7002 (1986), Agricultural food products – Layout for a standard method of sampling from a lot, ISO, Geneva.
32. IUPAC (2005), Terminology in soil sampling (IUPAC recommendation 2005), Prepared for publication by De Zorzi, P, Barbizzi, S, Belli, M, Ciceri, G, Fajgelj, A, Moore, D, Sansone, U. & Van der Perk, M., Pure and Applied Chemistry 77 (5) 827-841.
33. B. Magnusson, T. Näykki, H. Hovind, M. Krysell and E. Sahlin, Handbook for calculation of measurement uncertainty in environmental laboratories, Nordtest Report TR 537 (4th ed.) 2017. Available from www.nordtest.info.
34. B. Magnusson and U. Örnemark U. (eds.) Eurachem Guide: The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics, (2nd ed. 2014), ISBN 978-91-87461-59-0. Available from www.eurachem.org.

35. V. J. Barwick and E. Prichard (eds.), Eurachem Guide: Terminology in Analytical Measurement – Introduction to VIM 3 (2011). ISBN 978-0-948926-29-7. Available from www.eurachem.org.
36. EN 12823-1 (2014), Foodstuffs – determination of vitamin A by high performance liquid chromatography Part 1: Measurement of all-E-retinol and 13-Z-retinol, CEN.
37. AMC(2017), Analytical Methods Committee Technical Brief No. 78, Proficiency testing of sampling. Freely available from: https://pubs.rsc.org/en/content/articlelanding/2017/AY/C7AY90092A#fn.
38. M.H. Ramsey and S.L.R. Ellison (2015) Uncertainty Factor: an alternative way to express measurement uncertainty in chemical measurement. Accreditation and Quality Assurance. 20, 2,153-155. doi:10.1007/s00769-015-1115-6.

NT TECHN REPORT 604
Edition 2
Approved 2020-06

Return address:
Nordtest,
Slettetoften 9
DK-2630 Taastrup, DK

NORDTEST

NORDTEST, founded in 1973, is a Nordic institution acting as a joint body in the field competence and expertise in the field of harmonizing of norms and methods, a large Nordic net-work of experts, more than 650 recommended Nordic testing methods and 600 published technical reports.

www.nordtest.info