norden
Nordic Innovation Centre

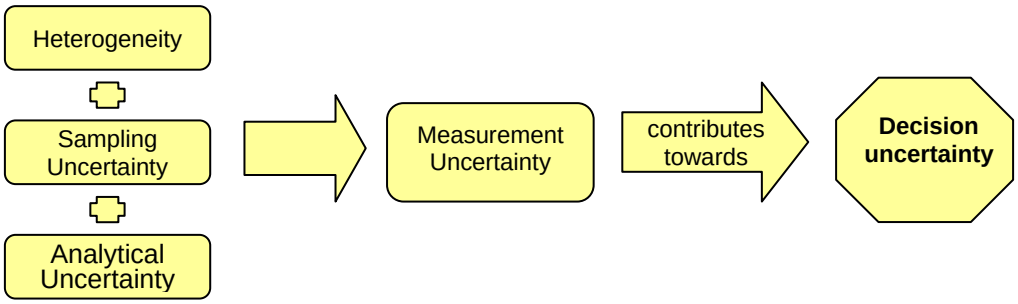
NT TECHNICAL REPORT

TR 604
Approved 2007-06

UNCERTAINTY FROM SAMPLING

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

BASED UPON THE EURACHEM INTERNATIONAL GUIDE
ESTIMATION OF MEASUREMENT UNCERTAINTY ARISING FROM SAMPLING



Christian Grøn, Jette Bjerre Hansen,
Bertil Magnusson, Astrid Nordbotten,
Mikael Krysell, Kirsten Jebjerg Andersen
and Ulla Lund

NORDTEST
a Nordic Innovation Centre brand

Nordic Innovation Centre
Stensberggata 25
NO-0170 OSLO
Norway

Telephone +47 47 61 44 00
Fax +47 22 56 55 65
info@nordicinnovation.net
www.nordicinnovation.net

ISSN 0283-7234

NT TECHN REPORT 604
Approved 2007-06

Authors:
Christian Grøn[1)], Jette Bjerre Hansen[1)], Bertil Magnusson[2)], Astrid Nordbotten[3)], Mikael Krysell[4)], Kirsten Jebjerg Andersen[4)] and Ulla Lund[4)]

Nordic Innovation Centre project number:
04130

Institution:
[1)] DHI
[2)] SP Technical Research Institute of Sweden
[3)] Food Safety Authority, Norway
[4)] Eurofins A/S

Title:

Uncertainty from sampling

A Nordtest Handbook for Sampling Planners on Sampling Quality Assurance and Uncertainty Estimation

Based upon the Eurachem International Guide *Estimation of Measurement Uncertainty Arising from Sampling*

Abstract:

The handbook provides practical guidance on sampling uncertainty estimation in the Nordtest TR handbook format. The handbook is an extract of and based upon the principles, methods and text of the international Eurachem Guide *Estimation of measurement uncertainty arising from sampling.* The Eurachem guide is more extensive and provides details on theory and additional examples. The basic reference for the text in this handbook is the above-mentioned Eurachem guide.

The overall purpose of this handbook has been to provide those working with investigations, monitoring and control that require sampling with a set of tools for calculation and control of the sampling uncertainty of their sampling procedure. Four examples illustrate the application of different 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.

It is the intention of the project group 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.

Technical Group:

ISSN:	**Language:**	**Pages:**
0283-7234	English	36 p. + 4 appendices

Key Words:
Sampling, uncertainty, food, ore, soil, waste, groundwater, waste water, validation, quality control

Distributed by:
Nordic Innovation Centre
Stensberggata 25
NO-0170 Oslo
Norway

Report Internet address:
info@nordicinnovation.net
www.nordicinnovation.net

Preface

The task of compiling and editing this handbook has been made possible by the financial support from Nordic Innovation Centre/Nordtest through project 04130. The work has been performed by a Nordic project group consisting of:

Christian Grøn, Jette Bjerre Hansen, DHI, Denmark
Bertil Magnusson, SP Technical Research Institute of Sweden
Mikael Krysell, Ulla O. Lund and Kirsten Jebjerg Andersen, Eurofins A/S, Denmark
Astrid Nordbotten, Food Safety Authority, Norway

The handbook provides practical guidance on sampling uncertainty estimation in the Nordtest TR handbook format. The handbook is an extract of and based upon the principles, methods and text of the international Eurachem Guide *Estimation of measurement uncertainty arising from sampling.* The Eurachem guide is more extensive and provides details on theory and additional examples. The latest information on the guide is available on the Eurachem website, www.eurachem.com. The basic reference for the text in this handbook is the above-mentioned Eurachem guide.

For valuable comments on the contents we thank Mike Ramsey and Paolo de Zorzi from the Eurachem working group.

The overall purpose of this handbook has been to provide those working with investigations, monitoring and control that require sampling with a set of tools for calculation and control of the sampling uncertainty of their sampling procedure. It is the intention of the project group 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. The project group hopes that this was achieved but if not, please recall that statistics generally seem more complicated than they are.

How to use the handbook

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

The background, theory and principles are described in text chapters, and worked out examples are given as appendices. The emphasis of the text chapters is upon simple explanations, with text boxes giving specific guidance and justification for the procedures, figures to illustrate the points made and example boxes to show the principles and practical doing of the calculations.

Chapter 1 is a description of the scope of the handbook

Chapters 2 and 3 provide the context of sampling as part of a measurement process with a purpose and a requirement for a defined quality.

Chapter 4 describes sources of error and uncertainty in sampling.

Chapter 5 describes how sampling validation and quality control can be designed.

Chapter 6 – the main chapter. Here we present the sampling designs and the statistical methods that enable calculation of sampling uncertainty.

In Chapter 7 and 8, short explanations and definitions of the terms used, as well as references and useful links are compiled.

The examples in the appendices illustrate the application of different 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.

Appendix	Matrix	Application	Sampling validation	QC	Design	Calculations
1	Groundwater	Monitoring	X	X	Double split	Relative range
2	Iron ore	Production	X	-	Double split	Range
3	Baby food	Surveillance	X	X	Empirical duplicate	ANOVA
4	Wastewater	Surveillance	-	-	Time series	Variography

"Spreadsheets and other calculation aids for the work can be found at www.samplersguide.com. Furthermore, a workshop presenting the principles, methods and tools was held in april 2007[*]. Presentations, cases and other materials from the workshop are available at www.samplersguide.com.

[*] Hansen et al., Accred Qual Assur (2007) 12:377–381

Contents

1 Scope

The aim of this handbook is to explain with detailed, practical examples some of the methods available for the estimation of uncertainty of sampling for *a predefined sampling procedure*. The examples provide assessments of whether the uncertainty obtained from a given sampling procedure is fit for purpose, i.e. fulfils pre-defined requirements. Furthermore, the handbook gives suggestions for design of sampling validation and quality control.

Although the examples given are for a limited selection of materials, the methods are generally applicable to most matrices.

2 Sampling in the measurement process

A complete measurement process, starting with primary sampling and ending in the analytical determination is shown in Figure 2-1. There are many intermediary steps, such as transportation and preservation of samples, not all of which are always present. Each step gives rise to a contribution to the uncertainty of measurement. The process steps of taking the sample(s) from the sampling target and doing the physical sample preparation (shaded boxes) are generally considered part of sampling and are done prior to delivering the sample at the door of the laboratory. A more extensive overview is given in the Eurachem guide [26].

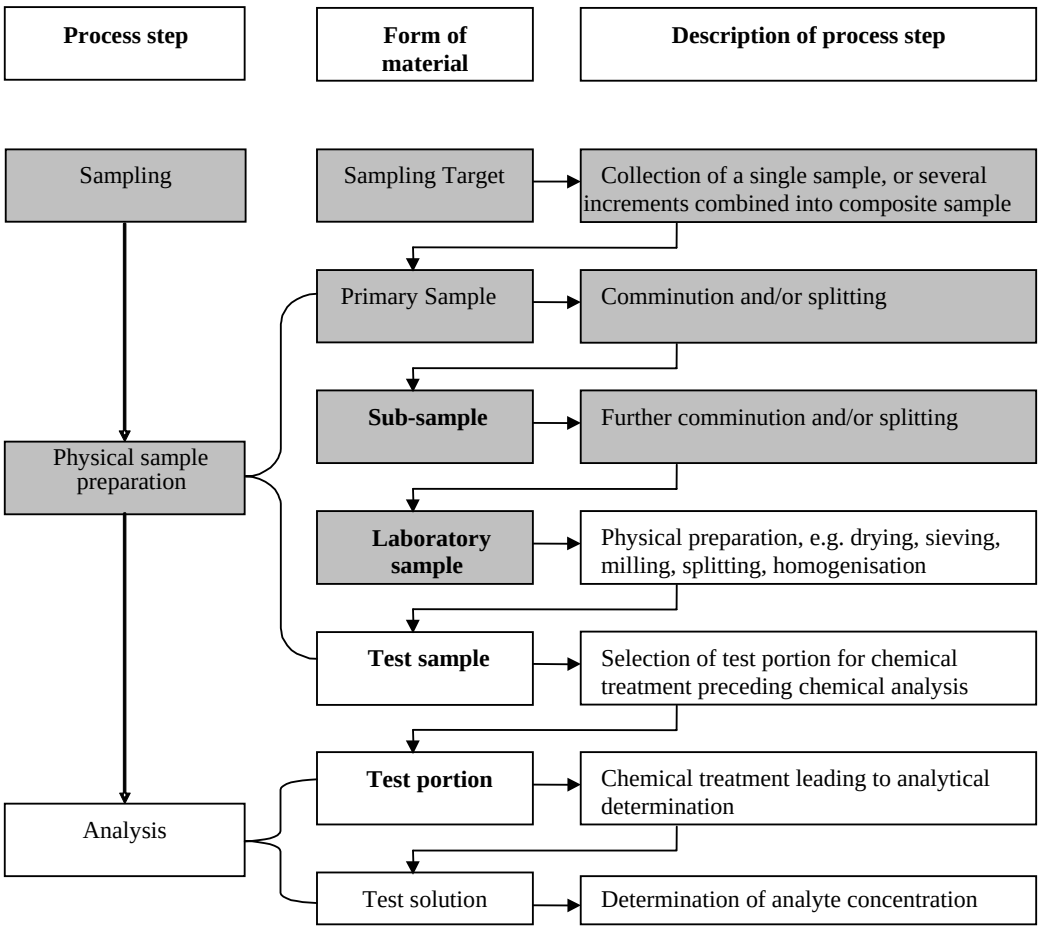


Figure 2-1 Schematic diagram of a typical measurement process including sampling, physical sample preparation (including transport) and analysis.

3 Sampling purpose and quality requirements

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

3.1 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 defined as being made up of two components: uncertainty derived from sampling a matrix and using those samples to represent the whole sampled mass; and 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.

3.2 Sampling target – what we shall measure

Based upon the purpose of the measurement, one has to define the sampling target, i.e. what 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. So far most measurement results have been presented to the end-user without any notion of the uncertainty, i.e. just as a figure, x. However, the trend is to present the result, X, as the measurement value, x, with the associated expanded uncertainty, U, see also Section 4.4:

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

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 *concentration in the bulk material, or sampling target*; the other *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 - e.g. total iron, Fe
- Unit and base for reporting – e.g. % 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.

3.3 Quality requirements - demands on uncertainty

Based on the evaluation of the purpose of the measurement it is necessary to set requirements for the uncertainty of the measurement. Examples of requirements are shown in *Box 3-1*.

Box 3-1 What kind of quality requirements could be useful?

The information we want can be qualitative:

- Sampling from a batch of tins with paint to determine the colour, e.g.: the paint is pink

or quantitative:

- Sampling from a shipment of footballs to find how many are OK, e.g.: 95% are OK, 5% are flat

For quantitative information, we also need to know how well the information describes the sampling target:

- The supplier of footballs may claim that you were just unlucky finding 5% flat ones in his shipment

This means that we need to know, in quantitative terms, how well the sample taken describes the sampling target. We can check all 100.000 footballs in the shipment, pretty expensive and time consuming, or we can take out 5 footballs at random, quick and cheap, but the supplier may not accept the result.

This means that we must beforehand decide how well the samples we take must describe the target that we are studying: can we accept to be 1% wrong in the quantitative information (i.e. 4-6% of the footballs are flat) or do we need to know with less or can we accept more uncertainty.

The sampling target that we are studying is not homogenous and the properties vary, and both sampling and analysis is associated with uncertainty. Any measurement is thus associated with an uncertainty that includes contributions from all of these factors: heterogeneity, sampling and analysis, and this measurement uncertainty interacts closely with decision making, Figure 3-1. 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.

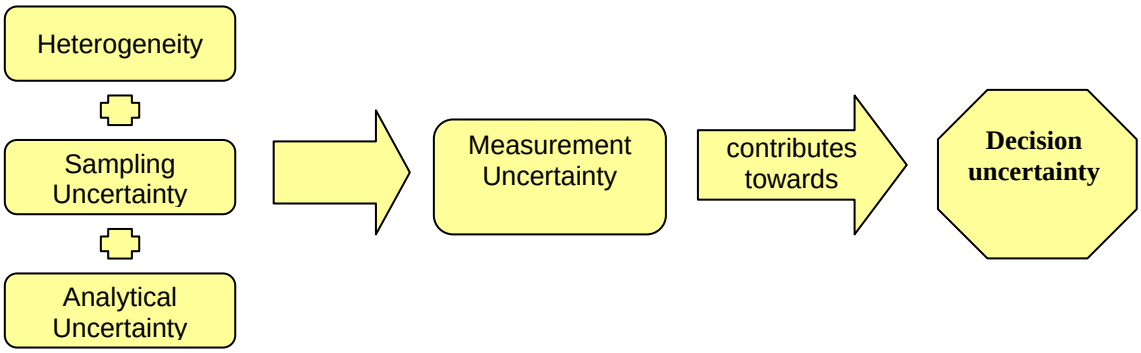


Figure 3-1 The uncertainty and decision chain.

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 upon this, quality requirements that are fit for purpose, *i.e.* are neither excessive (too expensive), nor insufficient (hopefully cheap), can be set.

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 as e.g.

- For the implementation of the EU Water Framework Directive: the limit of quantification (LOQ) must be 30% of the limit (the EQS) or lower and the expanded uncertainty must be 50% of the limit or lower [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 interlaboratory method study [4].
- For environmental control in Denmark: maximum standard deviation at low concentrations, maximum relative standard deviation at higher concentrations, maximum bias and maximum error, 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 and in many cases also for the uncertainty associated with between target variability. 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 3-2*.

Box 3-2 How can we set quality requirements from required certainty of decision?

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

The measurement result must differ from the MCL by at the least 2[1] times the standard uncertainty[2] in order to ascertain with a certainty of 95% that the mean is different from the MCL, see Section 4.4.

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

The analytical uncertainty was 4 (5% relative standard deviation), and that left room for a sampling uncertainty of 9.2 (11% relative standard deviation) while still maintaining the total uncertainty below the required value of 10 (12.5% relative standard deviation).

If the measurement had been 50, a standard uncertainty[2] of 25 could be accepted and with the same relative analytical uncertainty (5%), this would require a sampling uncertainty of not more than 24.9 or almost 50%. This would allow for taking significantly fewer sub-samples than with a measurement of 80 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 1.6 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 - see section 4.4.

4 Concepts of error and uncertainty in measurement

In this chapter 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 quantified and expressed.

4.1 Sources of uncertainty

The uncertainty arises from a variety of sources, and these can be categorized in different ways. The sources of uncertainty from analyses are well studied, but less focus has been upon uncertainty from sampling. In Table 4-1, some sources of uncertainty in sampling and sample preparation are listed.

Table 4-1 Some sources of uncertainty 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 thought to contribute to the uncertainty are presented in a *fish-bone diagram*. A typical example is presented in Figure 4-1.

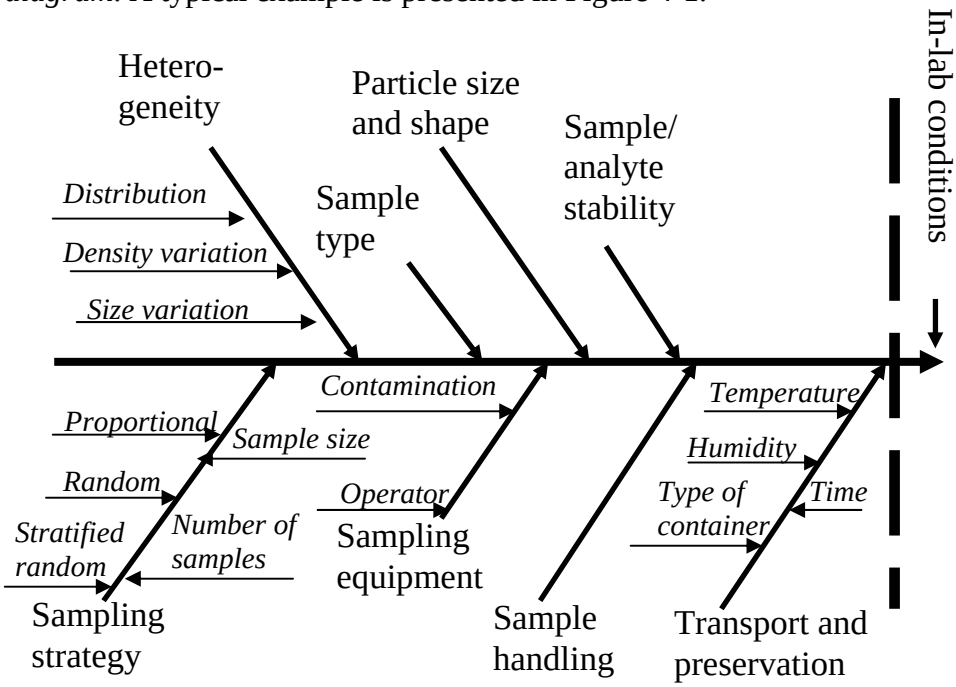


Figure 4-1 Cause-effect (fish-bone) diagram of possible sources contributing to the uncertainty.

An investigation of sources of uncertainty can be a useful help in identifying those steps in the measurement process that could contribute to measurement uncertainty and should be examined in case of excessive uncertainty.

Box 4-1 How can we use an investigation of the sources of uncertainty to pick the right point of action?

For a row of piles of contaminated soil as described in *Box 3-2*, the sampling standard uncertainty turned out to be too high to meet the quality requirement. The measurement was 80, but the measurement uncertainty was 25, not 10 as required. Using a replicate sampling design, Section 6.1, it was demonstrated that the excessive uncertainty was from the sampling (24.7), not from the analysis (4.0). A closer examination of the contaminant distribution in the piles 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 piles with different depths. Knowledge of the cause of the excessive measurement uncertainty made it possible to design a sampling method with a full cross section of a pile sampled for homogenization, splitting and subsampling that could provide the required lower sampling uncertainty of 10 [7].

4.2 Systematic and random effects

The terms systematic effects (bias) and random effects (precision) are familiar to most readers from the terminology used in studies of the uncertainty of laboratory analyses. 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 random effects only.

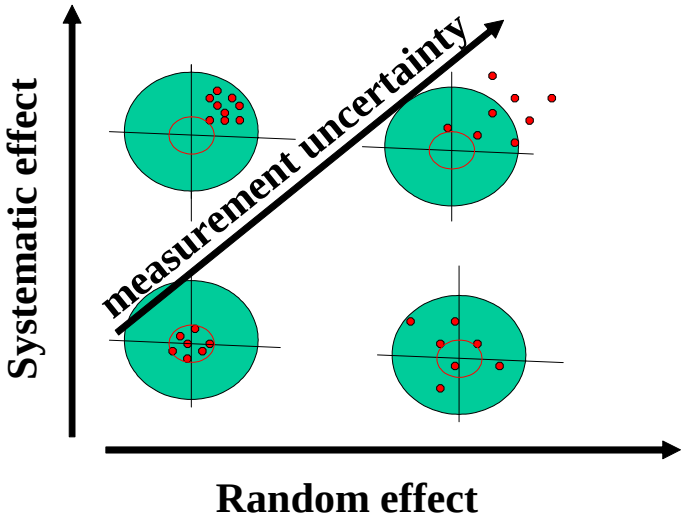


Figure 4-2 Random and systematic effects on analytical results and measurement uncertainty may be illustrated by the performance of someone practising at a target – the reference value or true value. Each point represents a reported analytical result.

The **systematic effects** in sampling are caused by the heterogeneity of the sampling target combined with an inability of the sampling method to properly reflect this heterogeneity. The heterogeneity can in turn be divided into the 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 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. The 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 preparation that match the sampling target and its properties 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 method, e.g. if different methods 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 the 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.

A general overview of the random and systematic effects in both analysis and sampling are given in Table 4-2. The effects and causes of systematic and random errors are further discussed in [8].

Table 4-2 Uncertainty contributions divided into random and systematic effects.

	Random *(precision)*	***Systematic*** *(bias)*
Analysis	Analytical variability - including sample splitting/preparation and handling (combined contribution of random effects)	Analytical bias (combined effect of bias sources)
Sampling	Sampling variability (dominated by heterogeneity and operator variations)	Sampling bias (combined effect of selection bias, operator bias etc.)

Box 4-2 How can knowledge on the type of sampling error help in designing the sampling?

The major source of uncertainty in contaminant measurements for a row of soil piles from *Box 3-2* was shown to be varying contaminant concentration with depth, i.e. a systematic effect. Sampling or sub-sampling of 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 sampling method was designed, as described.

4.3 Estimating sampling uncertainty

Both sampling and analysis contribute to measurement uncertainty. The uncertainty contribution due to physical sample preparation, transport, sample storage etc. is either included in the uncertainty of the sampling or of the analysis. The random part of the uncertainty is described by the standard deviation. The standard deviation of the measurement, $s_{measurement}$, is given by the following equation:

$$s^2_{measurement} = s^2_{sampling} + s^2_{analysis} \qquad \textit{Equation 2}$$

The basic tool to estimate the size of the random part of the measurement uncertainty is to repeat the measurement, i.e. to sample the same target and analyse the samples, corresponding to replicate analyses in the laboratory to quantify the analytical uncertainty, and to apply statistical calculations to the resulting analytical data.

The most practical way of estimating the sampling standard deviation, $s_{sampling}$, is the replicate method (see Section 6.1). The $s_{measurement}$ can be obtained from variation in results between samples and the analytical standard deviation, $s_{analysis}$, can be obtained from variation between analytical replicates. The $s_{sampling}$ can then be obtained by rearranging Equation 2:

$$s_{sampling} = \sqrt{s^2_{measurement} - s^2_{analysis}} \qquad \textit{Equation 3}$$

Box 4-3 Example of calculating the standard deviation for sampling

For the piles of contaminated soil described in *Box 3-2*, the analytical standard deviation was found to 4 and the measurement standard deviation to 10 in a split replicate study, see Section 6.1. The sampling standard deviation can then be found using Equation 3:

$$s_{sampling} = \sqrt{s^2_{measuremen\,t} - s^2_{analysis}} = \sqrt{10^2 - 4^2} = 9.2$$

It should also be noted that when applying the replicate design for determining the random part of the uncertainty, the standard deviation, the analytical contribution is from the analyses in the actual study only. Other contributions to the standard deviation of analyses such as effects due to using different instrument, day to day variation, using other operators might not be included.

Using the replicate design it should be noted that the systematic errors (bias) cannot be easily obtained, but some approaches to this are given in Table 4-3. The bias of analysis can be estimated by using certified reference materials (CRM) or participating in laboratory proficiency tests. As to the determination of the bias due to sampling – this is a more difficult task, as few sampling proficiency tests exist. Possible alternative approaches are:

- When a theoretical value is known, e.g. from production, and used as an estimate of the true value, see Appendix 3, or when sampling is performed on a reference sampling target [9]
- When two or more perform sampling and analysis, intersampler comparisons – e.g. when both producer and client perform sampling and analysis of the same batch of material, see Appendix 2
- When comparing results with those obtained using a detailed reference sampling method such as e.g. for sampling coating powders [10]

Table 4-3 Examples of tools for the estimation of uncertainty contributions.

	***Random** (precision)*	***Systematic** (bias)*
Analysis	Replicate analyses	Certified reference materials Laboratory proficiency test Reference analytical method
Sampling	Replicate samples	Reference sampling target Sampler proficiency test, Inter-method comparisons Known theoretical value of sampling target Reference sampling method

In this guide we focus on the random part of measurement uncertainty – the standard deviation *s*. In order to get the combined standard uncertainty, *u*, of sampling and analysis, we should include estimates of systematic effects see Table 4-3. Still, in the text we will use the terms analysis, sampling and measurement uncertainty even in cases where we consider only the random parts of the uncertainties. For the analytical part, the laboratory performing the analyses might have a realistic estimate of the analytical uncertainty where both random and systematic effects are considered. This important issue is further discussed in the examples in the appendices.

4.4 Uncertainty – *u* and *U*

The combined standard uncertainty, *u*, is calculated based on standard deviations of replicate measurements, *x*. It may, or may not, include contributions from systematic effects. In any case, as it is based on one single standard deviation, reporting a sampling uncertainty based directly on this value ($X = x \pm u$) will mean that the probability that the reported range contains the "true value" is only 67% (a 67% confidence interval). In most cases it is therefore more useful to the persons evaluating the data to use the expanded uncertainty, *U*.

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

$$U = 2 \cdot u \qquad \textit{Equation 4}$$

and the result, *x*, reported as $X = x \pm U$, see Equation 1, giving the range of the “true value”, *X*, with 95% confidence.

As stated above, the standard uncertainty *u* may in many cases be set equal to the standard deviation found from replicate measurements, *s*.

This approach to reporting measurements with their associated uncertainty will give a 95% confidence interval and thus the interval around a measurement that will include the “true value” with 95% certainty. The uncertainty should ideally include all steps from sampling and analysis from both random and systematic effects.

5 Principles of quality assurance in sampling

In this chapter guidance is provided upon quality assurance of sampling, including the required competence, validation and quality control of sampling methods, and documentation of sampling.

5.1 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 a medical doctor, 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 method and the sampling equipment
- Competence about the sample from 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 with the person or even institution in charge. Therefore, it is the obligation for the responsible person or institution to acquire the external competencies required to cover the entire field.

Box 5-1 How can the required competence suite be established?

The piles of contaminated soils mentioned in *Box 4-3* were situated at a soil remediation facility. A consultant was trusted with the task of making a risk assessment of the piles 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 methods, 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 competencies documented by e.g. accreditation of the organisation to perform the sampling methods [11] or certification of persons for environmental sampling [12,13].

5.2 Principles for sampling validation and quality control

Once the quality requirements have been set, the next step is to set the sampling and analytical uncertainty needed to meet these requirements. To evaluate the sampling and analytical uncertainty two tools can be chosen and combined here: 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 waste water). 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 5-1.

Table 5-1 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 [14]. Specific suggestions for sampling quality control are given for some environmental sampling matrices in [13]. The methods used in validation are, with some simplification, applicable to quality control, but quality control is in most cases less extensive than the 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 target with the uncertainty established at the validation.

The focus of quality control is almost exclusively the random aspect, whereas the systematic effects are difficult to address in validation and almost impossible in quality control. The flow in designing validation and quality control is shown in *Box 5-2*.

The principal tool for validation is replicate measurements, mostly in a split level design; see Section 6.1 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 upon an analysis of the contributions to uncertainty from different sources, see e.g. Section 4.1. In Appendix 3, an example is given where the validation of a measurement process (control of batches of baby porridge for vitamin A content) 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, see Section 6.1.

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 5-2 How can we design a validation and quality control programme?

The design of validation and quality control programmes is demonstrated in Appendices 1 and 3 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 do a validation study with at least 8 duplicate samplings with separate analysis of each sample and with the samplings varying in space (different points within the target) or in time (different sampling times) depending upon 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 Chapter 6, 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, see Section 6.1. If an early warning of a measurement process out of control is required, control charts can be used.

Range control charts [15] are constructed from e.g. duplicate measurements where the difference between the results:

$$D = |x_1 - x_2| \qquad \textit{Equation 5}$$

is calculated and the standard deviation of measurement

$$s_{measurement} = \frac{D}{1.128} \qquad \textit{Equation 6}$$

A one-sided range control chart can be constructed with a control limit of $2.83 \cdot s_{measurement}$ (yellow/light grey, not exceeded in 95% of control result) and an action limit of $3.69 \cdot s_{measurement}$ (red/black, not exceeded in 99% of control). For details on construction of the chart, see *Box 5-3*.

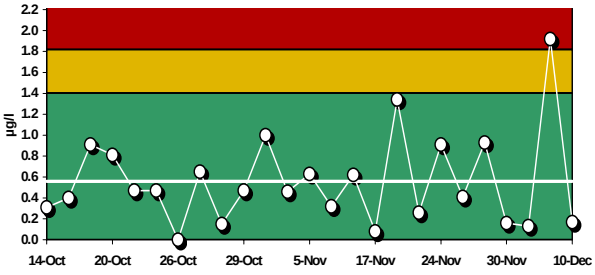


Figure 5-1: Example of a range control chart for quality control of sampling.

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.

Box 5-3 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 many standard laboratory software packages. The basic steps are [15]:

- Calculate the measurement standard deviation, $s_{measurement}$, from the validation study
- Set the baseline to zero
- Set the central line, CL, to $1.128 \cdot s_{measurement}$
- Calculate the warning limit, WL, as $2.83 \cdot s_{measurement}$ and plot this value as a horizontal line in the chart
- Calculate the action limit, AL, as $3.69 \cdot s_{measurement}$ 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 $D = |x_1 - x_2|$
- 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

Appendix 3 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 $d = \frac{|x_1 - x_2|}{\bar{x}}$.

5.3 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 5-2.

Table 5-2 Summary of sampling documentation [16,17,18].

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 specific and detailed description of the operations used for sampling after a defined principle and with defined equipment. May be the operational implementation of a sampling method.
Sampling field report	The detailed notes on the sampling details as noted 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 results including target definition, reference to applied method and procedure, relevant notes from field and chain of custody report and uncertainty.

Documentation should include at least:

- Written sampling procedures based upon defined sampling methods
- Sampling field report
- Sampling reports including the uncertainty of sampling

Specific, 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 provides the summary results of the sampling and, including the sampling uncertainty, allows the end user to evaluate the sampling quality against the defined quality requirements. The sampling report may be part of the measurement (analytical) report.

Details on documentation of sampling can be found in [13,19].

6 Methods and calculations

In this chapter we present sampling uncertainty calculations. The random sampling uncertainty is based upon replication of the measurement procedure or parts of it by splitting the samples (replicate design) or replication of the measurement procedure in time or space (variography).

6.1 The replicate design

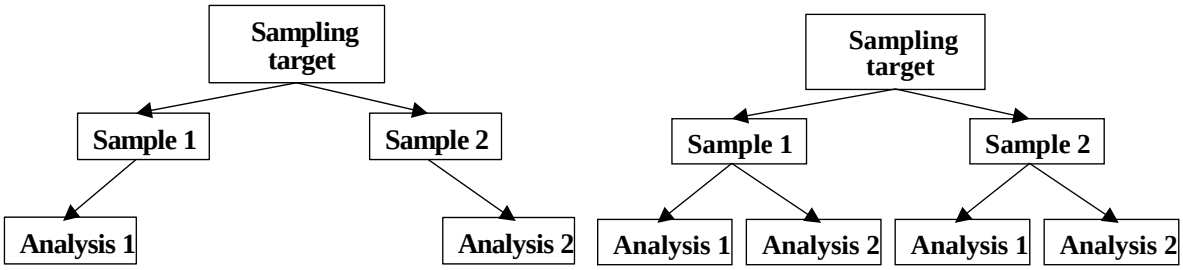


Figure 6-1 The principles of the replicate design with one (left) and two (right) split levels.

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 to estimate the random measurement error, preferentially at least 8 times for each calculation. The replication can be done in one step, as e.g. the sampling, but can also be done in other, critical, steps where information on the uncertainty is required, as e.g. the analysis or the sample preparation, see also Section 5.2. The replicate design is illustrated in Figure 6-1 for a design with one and two split levels. The use of the symbols in the replicate design is shown in Table 6-1. 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 6-1 Symbols used in the calculations using the replicate design x_{ijk} , with two split levels for three sampling targets (i). For each replication of a target, two samples (j) are taken and each sample is split into two subsamples (k) for analysis. The table shows the calculation of a mean range value for analysis, $\overline{D}_{analysis}$ and for measurement, $\overline{D}_{measurement}$.

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

With the design shown in Figure 6-1, the contributions from sampling and analytical random errors to the total uncertainty can be estimated. The design can be adapted/extended to provide information on other contributions such as e.g. sub-sampling, preservation, transportation and storage of samples.

6.2 Range statistics

Range statistics are used to calculate the standard deviation, *s*, or the relative standard deviation, *RSD*, for the steps in the split design. Normal distribution of the data is assumed. 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 *RSD* 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 the relative standard deviation is constant over the measuring range. Based upon duplicate data, this has been suggested to be the case for most environmental and geochemical purposes at least with concentrations above ≈100 times the limit of detection [20].

The calculation of the (relative) standard deviation from the mean (relative) differences is based upon a statistical analysis of the relation between standard deviation and differences, and the factor applied depends upon the replication chosen, e.g. duplicate, triplicate [15]. Similar estimates would be obtained if the standard deviations were calculated for each set of duplicates and combined as variances.

6.2.1 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:

$$D_i = |x_{i1} - x_{i2}| \qquad \textit{Equation 7}$$

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

$$\bar{x}_i = \frac{x_{i1} + x_{i2}}{2} \qquad \textit{Equation 8}$$

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

$$d_i = \frac{D_i}{\bar{x}_i} \qquad \textit{Equation 9}$$

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

$$\bar{d} = \frac{\Sigma\, d_i}{n} \qquad \textit{Equation 10}$$

The relative standard deviation, *RSD*, for measurement is calculated using a statistical constant [21] of 1.128 (when analysing duplicates):

$$RSD = \frac{\bar{d} * 100}{1.128}\% \qquad \textit{Equation 11}$$

The standard deviation, s_{x_0}, at a given concentration, x_0, can be estimated from:

$$s_{x_0} = \frac{RSD * x_0}{100} \qquad \textit{Equation 12}$$

An example of the calculations is shown in details in *Box 6-1*.

Box 6-1 Calculation example demonstrating the use of relative range statistics for calculating relative standard deviation from duplicates (one split relative range statistics)

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

$\mathbf{x}_{i1}$	$\mathbf{x}_{i2}$	$D_i = \lvert x_{i1} - x_{i2} \rvert$	$\bar{x}_i = (x_{i1} + x_{i2})/2$	$d_i = D_i / \bar{x}_i$
20	2	18	11	1.64
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	224	170	139	1.22
442	325	117	384	0.31
765	755	10	760	0.01
32	516	484	274	1.77
650	15	635	333	1.91

Mean relative range	Relative measurement standard deviation	*Standard deviation at a level of 200 mg/kg*
$\bar{d} = \Sigma\, d_i / n = 0.93$	$RSD = 100 \cdot \bar{d} / 1.128 = 82\%$	$s_{200} = RSD * 200/100 = 164$ *mg/kg*

The application of relative range statistics is demonstrated in Appendix 1 for validation and quality control of groundwater sampling.

6.2.2 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, see Table 6-1. Appendix 3 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 shown in Table 6-1 requires that all measurements be within a range where the standard deviation is close to constant. In *Box 6-2*, the calculation of standard deviation for the different split steps is shown with the data from Appendix 3.

Box 6-2 Calculation example demonstrating the use of range statistics for calculating standard deviation from duplicate samples and duplicate analyses (two split range statistics)

Duplicate samples were taken from 10 batches of baby porridge and analysed for vitamin A (µg/100 g) in duplicate and calculations done as follows (see Appendix 3 for more details):

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
$\bar{D}_{i1} = \frac{\sum D_{i1}}{n}$		36.5		$\bar{D}_{i2} = \frac{\sum D_{i2}}{n}$		30.7	$\bar{D} = \frac{\sum D_i}{n}$	32.1

Mean range of analysis

$$\bar{D}_{analysis} = \frac{\bar{D}_{i1} + \bar{D}_{i2}}{2} = 33.6$$

Standard deviation of analysis

$$s_{analysis} = \frac{\bar{D}_{analysis}}{1.128} = 29.8$$

Mean range of measurement

$$\bar{D} = 32.1$$

Standard deviation of measurement based on duplicate analysis

$$s_{measurement} = \frac{\bar{D}}{1.128} = 28.5$$

Standard deviation of sampling $\quad s_{sampling} = \sqrt{s_{measurement}^2 - \left(\frac{s_{analysis}}{\sqrt{2}}\right)^2} = 19.1$

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 formula above – standard error of the mean.

6.3 ANOVA

Using the split replicate design the standard deviations can also be estimated by applying one-way analysis of variances (ANOVA). The variance, *V*, is defined as the square of the standard deviation:

$$V = s^2 \qquad \textit{Equation 13}$$

The values of $s_{sampling}$ and $s_{analysis}$ are thus obtained from the corresponding variances, $V_{sampling}$ and $V_{analysis}$, estimated with the ANOVA. A suitable experimental design is shown in Figure 6-1. The source of the variation considered in this design will be the *between analyses variance, $V_{analysis}$*, and the *between sample variance, $V_{sampling}$*.

The ANOVA calculations can be performed using standard spreadsheet data analysis functions or dedicated software. There is often a small proportion (i.e. <10%) of outlying values in the frequency distributions of the analytical, within-sample and

between-sample variability. This requires the use of some method of down-weighting the effect of the outlying values on classical ANOVA, such as the use of robust statistical methods 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 there may also be the possibility to use *robust ANOVA*.

In this section we will describe in detail how the ANOVA calculations are performed. It should be emphasized that ANOVA calculations are more complicated than range statistics and more detailed information may be required, see e.g. [22]. Still, in many cases the ANOVA can be performed using spreadsheets or dedicated software with reasonable efforts. 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.

The type of ANOVA used here is one-way ANOVA, meaning that the calculation deals with one independent variable and one dependent variable. It is assumed that the analyses of each group have a normal distribution and that each group has the same distribution.

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 6.2. 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 according to:

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

See Table 6-1 for an explanation of the symbols.

Then, for each of the two samples the differences *D* 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);

$$| x_{i11} - \bar{x}_{i1} | = | x_{i12} - \bar{x}_{i1} | = D_{i1(\bar{x})}$$ *Equation 15*

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

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

The sum of squares of differences within groups, $SS_{E\text{-analysis}}$, is calculated by summation of the sum of the squares of all the samples:

$$SS_{E-analysis} = 2 * \sum_{i=1}^{10} [D^2_{i1(\bar{x})} + D^2_{i2(\bar{x})}]$$ *Equation 17*

The degrees of freedom, $df_{analysis}$, is calculated from

$$df_{analysis} = i \cdot j \cdot k - i \cdot j$$ *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, $V_{analysis}$, is then calculated as:

$$V_{analysis} = \frac{SS_{E-analysis}}{df_{analysis}}$$ *Equation 19*

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

$$s_{analysis} = \sqrt{V_{analysis}}$$ *Equation 20*

$$RSD_{analysis} = \frac{s_{analysis}}{\overline{X}} * 100\%$$ *Equation 21*

where $\overline{X}$ is the mean of all results.

In Box 6-3, the calculations are demonstrated for the same raw data used in Section 6.2.2 and in Appendix 3.

Box 6-3 Calculation example demonstrating the use of ANOVA for calculating standard deviation of analysis from duplicate samples and duplicate analyses (two split replicate design)

Duplicate 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 Appendix 3 for more details):

S1A1	S1A2	S2A1	S2A2	S1	S2	S1	S2
x_{i11}	x_{i12}	x_{i21}	x_{i22}	$\overline{x}_{i1}$	$\overline{x}_{i2}$	$2 * D^2_{i1(\overline{x})}$	$2 * D^2_{i2(\overline{x})}$
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
$\overline{X} = 347.9$						$SS_{E\text{-}analysis} = 2 * \sum_{i=1}^{10} [D^2_{i1(\overline{x})} + D^2_{i2(\overline{x})}] = 16595$	

$df_{analysis} = (i*j-i) = (10*2*2)-(10*2) = 20$	$V_{analysis} = SS_{E\text{-}analysis}/df_{analysis} = 16595/20 = 829.75$
$s_{analysis} = \sqrt{V_{analysis}} = \sqrt{829.75} = 28.8$	$RSD_{analysis} = \frac{s_{analysis}}{\overline{X}} * 100\% = \frac{28.8}{347.9} * 100\% = 8.28\%$

Now, the variance of measurement can be estimated. The mean value of each batch, *i*, (two samples - 4 analyses) is calculated as

$$\overline{X}_i = \frac{\overline{x}_{i1} + \overline{x}_{i2}}{2}$$ *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 $(D_{i(\overline{x})})^2$ between the mean value of the batch and the mean value of each sample in the batch is calculated according to:

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

The sum of squares of sampling $SS_{sampling}$ is calculated according to:

$$SS_{sampling} = \sum_{i=1}^{10}\left[\left(\frac{x_{i11}+x_{i12}}{2} - \overline{X}_i\right)^2 + \left(\frac{x_{i11}+x_{i12}}{2} - \overline{X}_i\right)^2 + \left(\frac{x_{i21}+x_{i22}}{2} - \overline{X}_i\right)^2 + \left(\frac{x_{i21}+x_{i22}}{2} - \overline{X}_i\right)^2\right] \qquad \textit{Equation 24}$$

$$= \sum_{i=1}^{10}(2*D_{i(\bar{x})}{}^2 + 2*D_{i(\bar{x})}{}^2) = 4*\sum_{i=1}^{10} D_{i(\bar{x})}{}^2$$

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

$$df_{sampling} = i \cdot j - i \qquad \textit{Equation 25}$$

The variance of sampling, $V_{sampling}$ is then calculated according to:

$$V_{sampling} = (SS_{sampling}/df_{Sampling} - SS_{analysis}/df_{analysis})/2 \qquad \textit{Equation 26}$$

The standard deviation, $s_{sampling}$, and relative standard deviation, $RSD_{sampling}$, of sampling are calculated

$$\mathrm{s}_{\mathrm{sampling}} = \sqrt{V_{sampling}} \qquad \textit{Equation 27}$$

If $V_{sampling} < 0$ then $s_{sampling}$ is conventionally set to zero.

$$RSD_{sampling} = (\frac{s_{sampling}}{\overline{X}}) * 100\% \qquad \textit{Equation 28}$$

In *Box 6-4* the calculations are demonstrated for the same raw data used in Section 6.2.2 and in Appendix 3.

Box 6-4 Calculation example demonstrating the use of ANOVA for calculating standard deviation of sampling 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 Appendix 3 for more details):

S1A1	**S1A2**	**S2A1**	**S2A2**	**S1**	**S2**	$\overline{X}_i =$	$(D_{i(\bar{x})})^2 =$
$\mathbf{x_{i11}}$	$\mathbf{x_{i12}}$	$\mathbf{x_{i21}}$	$\mathbf{x_{i22}}$	$\bar{x}_{i1}$	$\bar{x}_{i2}$	$\frac{\bar{x}_{i1}+\bar{x}_{i2}}{2}$	$\left(\overline{X}_i-\bar{x}_{i1}\right)^2=\left(\overline{X}_i-\bar{x}_{i2}\right)^2$
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
$\overline{X}=347.9$				$SS_{sampling}=4*\sum_{i=1}^{10}\left[\left(D_{i(\bar{x})}\right)^2\right]=14231$		$SS_{E\text{-}analysis}$ = 16595 (see *Box 6-3*) $df_{analysis}$ = 20 (see *Box 6-3*)	

$df_{sampling}$ = (*i***j*-*i*)=(10*2-10)= 10	$V_{sampling}$ = ($SS_{sampling}/df_{sampling} - SS_{analysis}/df_{analysis}$)/2 = (14231/10-16595/20)/2 = 296.675
$s_{sampling}=\sqrt{V_{sampling}}=17.244$	$RSD_{sampling}=\frac{s_{sampling}}{\overline{X}}*100\%=\frac{17.224}{347.9}*100\%=4.95\%$

The calculation procedures for *robust ANOVA* are not explained and reference is made to available software packages.

Selection of calculation method

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

- The statistical competence of the person in charge
- The complexity of the design behind the data
- The access to calculation tools such as spreadsheets and software packages

As a help in selecting the appropriate method, the results obtained using 4 calculation methods, range (single and double split), ANOVA, and *robust ANOVA* for the same dataset (*Box 6-2*) are shown in Table 6-2. Robust ANOVA calculations were done with the software package ROBAN [23], based on a published progamme [24].

Table 6-2 Examples of results calculated using range statistics and ANOVA on data in Box 6-2 – vitamin A in baby food.

	$s_{analysis}$	$RSD_{analysis}$	$s_{sampling}$	$RSD_{sampling}$	$s_{measurement}$
	µg/100 g	%	µg/100 g	%	
Range – single split	-	-	-	-	42
Range – double split	30	8.6	19	5.5	35
ANOVA	29	8.3	17	5.0	34
Robust ANOVA	31	8.8	21	6.1	37

Evidently, the differences in statistical estimates obtained with the 4 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 many and/or extreme outliers.

6.4 Variography

Variography is used to determine variations 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 is, together with knowledge about the analytical error, a tool for identifying and quantifying (part of) the uncertainty components of the random sampling errors, 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 sample 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 sampling uncertainty calculations, the most useful feature of a variographic experiment is the ability to estimate the variability between two samples taken with 0 time difference 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 samplings 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 6-2.

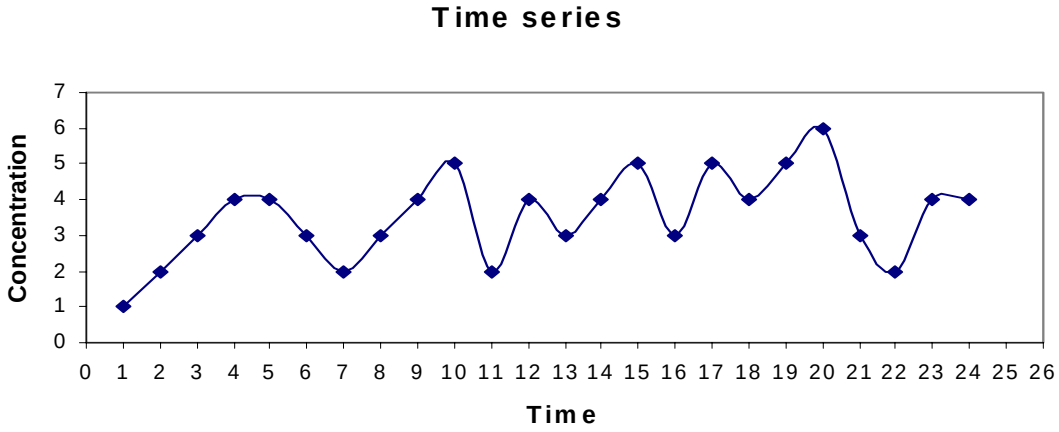


Figure 6-2 Example of a time series where the measured concentration is plotted against time [25].

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:

$$V(j) = \frac{\left[\sum_{i=1}^{n-j}(x_{i+j} - x_i)^2\right]}{\left[2\cdot(n-j)\cdot(\bar{x})^2\right]} \qquad \textit{Equation 29}$$

where x_i are the measured concentrations at the times *i*, *j* is the 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 average concentration over the time series. A variogram based on the time series given in Figure 6-2 is shown in Figure 6-3.

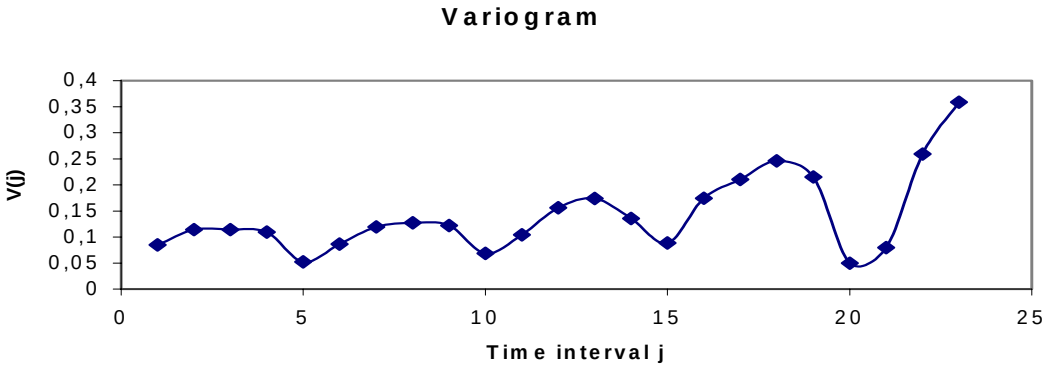


Figure 6-3 Variogram for the time series in Figure 6-2. On the x-axis are given the different time intervals, j, in this case 1 to 23, that have been used to calculate the variances, V(j) on the y-axis [25].

From the variogram it is obvious that the process in question has a periodic cycle consisting of 5 one hour time 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 e.g. sampling intervals to have information about cyclic variations of the sampling target. Note that this means that the points to the right in the diagram 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 hours measurement minus the 1 hours measurement). In many cases the variances in the last 5-10 points of the diagram may 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:

$$s^2(0) = V(0) \cdot \bar{x}^2 \qquad \textit{Equation 30}$$

then recalculated to a coefficient of variation, *RSD*:

$$RSD = 100 * \frac{s(0)}{\bar{x}} \qquad \textit{Equation 31}$$

s(0) is a representation of $s_{measurement}$ 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 sources of uncertainty 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: $s^2_{measurement} = s^2_{sampling} + s^2_{analysis}$.

V(0) will represent the $s_{measurement}$. 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: $s_{sampling} = \sqrt{s^2_{measurement} - s^2_{analysis}}$.

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 $s_{sampling}$ thus determined gives the uncertainty of sampling 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 details in Appendix 4.

7 Terminology and symbols

Symbols used in this guide

s	Standard deviation
RSD	Relative standard deviation in %
x	Measured value
x_{ijk}	Measured vale from target or batch (i), sample (j) and split (k)
$\bar{x}$	Mean value
X	Measurand
R_w	Within-laboratory reproducibility
D	Range from difference; $D = \lvert x_1 - x_2 \rvert$
d	Relative range from difference
$D_{\bar{x}}$	Range from mean value; $D_{\bar{x}}$ value – mean value
u	Standard uncertainty
U	Expanded uncertainty at 95% confidence level, $U = 2 \cdot u$

The terminology below is mainly based on [26]

Analyte Substance or parameter subject to measurement [15]

Bias The difference between the expectation of the test result and an accepted reference value
Note: Bias is a measure of the total systematic error as contrasted to random error. There may be one or more systematic error components contributing to the bias. A larger systematic difference from the accepted reference value is reflected by a larger bias value.
ISO 3534-1: 3.13 (1993)[27]

Composite sample (Also average and aggregate) Two or more increments/sub-samples mixed together in appropriate portions, either discretely or continuously (blended composite sample), from which the average value of a desired characteristic may be obtained.
ISO 11074-2: 3.10 (1998) [17], AMC (2005) [28]

Duplicate (Replicate) sample One of the two (or more*) samples or sub-samples obtained separately at the same time by the same sampling procedure or sub-sampling procedure
*for replicate sample
Note: each duplicate sample is obtained from a separate 'sampling point' within the 'sampling location'
Adapted from ISO 11074-2: 2.14 (1998) [17]; ISO 11074-2:1998 was formally adapted from ISO 3534-1 (1993); AMC (2005) [28]

Error of result	The test result minus the accepted reference value (of the characteristic) *Note: Error is the sum of random errors and systematic errors* ISO 3534-1: 3.8 (1993)[27]
Estimation	The term is used in statistics for calculations, where only an unlimited number of data points would provide accurate calculated values. The estimated value is calculated with less data and is thus not accurate but an estimate. Most non-statisticians would use the term calculation instead of estimation.
Fitness for Purpose	The degree to which data produced by a measurement process enables a user to make technically and administratively correct decisions for a stated purpose *Note: as defined for analytical science* Thompson and Ramsey, 1995 [29]
Homogeneity, heterogeneity	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 (1990) [30]; ISO 11074-2: 1.6 (1998) [17]
Increment	Individual portion of material collected by a single operation of a sampling device IUPAC (1990) [30], AMC (2005) [28]
Laboratory Sample	Sample as prepared for sending to the laboratory and intended for inspection or testing. ISO 78-2 (1999) [31], adopted by CAC
Measurand	Particular quantity subject to measurement ISO-GUM (1993) [32] *Authors' note:The specification of measurand regarding sampling target, analyte, unit and base for reporting is discussed in Section 3.2.*
Precision	The closeness of agreement between independent test results obtained under stipulated conditions *Note 1. Precision depends only on the distribution of random errors and does not relate to the true value or the specified value* *Note 2. The measure of precision usually is expressed in terms of imprecision and computed as a standard deviation of the test results. Less precision is reflected by a larger standard deviation* *Note 3. "Independent test results" means results obtained in a manner not influenced by any previous result on the same or similar test object. Quantitative measures of precision depend critically on the stipulated conditions. Repeatability and reproducibility conditions are particular sets of extreme stipulated conditions.* ISO 3534-1: 3.14 (1993) [27]

Primary sample	The collection of one or more increments or units initially taken from a population *Note: The term primary, in this case, does not refer to the quality of the sample, rather the fact that the sample was taken during the earliest stage of measurement.* IUPAC (1990) [30], AMC (2005) [28]
Random error of result	A component of the error which, in the course of a number of test results for the same characteristic, remains constant or varies in an unpredictable way *Note: It is not possible to correct for random error* ISO 3534-1: 1993 (3.9) [27]
Random sample	A sample of n sampling units taken from a population in such a way that each of the possible combinations of n sampling units has a particular probability of being taken ISO 3534-1: 4.8 (1993) [27]
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: A.34 (1986) [33]
Reference sampling	Characterisation of an area, using a single sampling device and a single laboratory, to a detail allowing the set-up of a distribution model in order to predict element concentrations, with known uncertainty, at any sampling point IUPAC (2005) [34]
Reference sampling target (RST)	The analogue in sampling of an 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 (1995) [29]
Representative sample	Sample resulting from a sampling plan that can be expected to reflect adequately the properties of interest in the parent population IUPAC (1990) [30]; ISO 11074-2: 1.9 (1998) [17], AMC (2005) [28]
Sample	A portion of material selected from a larger quantity of material IUPAC (1990) [30]; ISO 11074-2 (1998) [17], AMC (2005) [28]

Sample preparation The set of material operations (such as reduction of sizes, mixing, dividing, etc.) that may be necessary to transform an aggregated or bulk sample into a laboratory or test sample
Note: The sample preparation should not, as far as possible, modify the ability of the sample to represent the population from which it was taken
Adapted from ISO 3534-1: 4.30 (1993) [27]

Sample pre-treatment Collective noun for all procedures used for conditioning a sample to a defined state which allows subsequent examination or analysis or long-term storage
Adapted from ISO 11074-2: 6.1 (1998) [17]

Sample size Number of items or the quantity of material constituting a sample
ISO 11074-2: 4.26 (1998) [17]; ISO 7002: A.40 (1986) [33].

Sampler Person (or group of persons) carrying out the sampling procedures at the sampling point
Note: The term 'sampler' does not refer to the instrument used for sampling, i.e. the 'sampling device'
Adapted from ISO 11074-2 (1998) [17]

Sampling Process of drawing or constituting a sample
Note: For the purpose of soil investigation 'sampling' also relates to the selection of locations for the purpose of in situ testing carried out in the field without removal of material (from ISO 1998)
ISO 11074-2 (1998) [17]; ISO 3534-1 (1993) [27]

Sampling Bias The part of the measurement bias attributable to the sampling
AMC (2005) [28]

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
IUPAC (1990) [30]; ISO 11074-2 (1998) [17], AMC (2005) [28]

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 (2005) [28]

Sampling procedure Operational requirements and/or instructions relating to the use of a particular sampling plan; i.e. the planned method of selection, withdrawal and preparation of sample(s) from a lot to yield knowledge of the characteristic(s) of the lot
ISO 3534-1: 4.5 (1993) [27]; ISO 11704-2 (in part) [17], AMC (2005) [28]

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 (2005) [28]

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: 4.8 (1993) [27]

Sub-sampling (Sample division) Process of selection one or more sub-samples from a sample of a population
ISO 11074-2 (1998) [17]

Systematic error of result A component of the error which, in the course of a number of test results for the same characteristic, remains constant or varies in a predictable way.
Note: systematic effects and their causes may be known or unknown
ISO 3534-1: 1993 (3.10) [27]

Systematic sampling Sampling by some systematic method
ISO 3534-1: 4.15 (1993) [27]; ISO 11074-2 (1998)[17]

Test portion Quantity of material, of proper size for measurement of the concentration or other property of interest, removed from the test sample
IUPAC (1990) [30]; ISO 11074-2: 3.17 (1998) [17], AMC (2005) [28]

Test sample Sample, prepared from the laboratory sample, from which the test portions are removed for testing or analysis
IUPAC (1990) [30]; ISO 11074-2: 3.16 (1998) [17], AMC (2005) [28]

Trueness	The closeness of agreement between the average value obtained from a large series of test results and an accepted reference value. *Note 1. The measure of trueness is usually expressed in terms of bias,* *Note 2. The trueness has been referred to as "accuracy of the mean". This usage is not recommended* ISO 3534-1: 1993 (3.12) [27]
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.* *Note 4. (added) If measurand is defined in terms of the quantity within the sampling target, then uncertainty from sampling is included within uncertainty of measurement* ISO GUM: B.2.18 (1993) [32]
Uncertainty from sampling	The part of the total measurement uncertainty attributable to sampling IUPAC (2005) [34]

8 References and links

[1] - Draft council decision Implementing Directive 2000/60/EC Concerning Minimum Performance Criteria For Analytical Methods Used For Chemical Monitoring And The Quality Of Analytical Results.

[2] - Horwitz W. (1980) Quality Assurance in the Analysis of Foods for Trace Constituents, *Journal of the AOAC* 63:6, 1344-1354.

[3] - Thompson M. (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 nr 1353 af 11/12/2006 Bekendtgørelse om kvalitetskrav til miljømålinger udført af akkrediterede laboratorier, certificerede personer m.v.

[6] - Water Research Centre (1978), Manual on Analytical Quality Control for the Water Industry, TR66.

[7] - Kim Esbensen (2006), Aalborg University Esbjerg, proposal for sampling method.

[8] - Patricia L. Smith (2001) A primer for sampling solids, liquids, and gases. ASA, SIAM.

[9] - Barbizzi, S., de Zorzi, P., Belli, M.; Pati, A., Sansone, U., Stellato, L., Barbina, M., Deluisa, A., Menegon, S. & Coletti, V. (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 (2005), General requirements for the compentence of testing and calibration laboratorier, CEN.

[12] - EN ISO/IEC 17024 (2003), Conformity assessment – Generel requirements for bodies operating certification of persons, CEN.

[13] - Nordtest Sampler Certification Scheme Handbook, Version 1-0 (2005), Nordic Innovation Centre, www.nordicinnovation.net/nordtest.

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

[15] - Magnusson, B., Hovind, H., Krysell, M. & I. Mäkinen (2005), Internal Quality Control, Handbook for Chemical Laboratories, NORDTEST report TR 569, 2nd Ed.. Internet version 2005, .www.nordicinnovation.net/nordtest.cfm under link Rapporter, accessed 2006-08-30.

[16] - EN 45020 (1996), Standardization and related activities – General vocabulary.

[17] - ISO 11074-2 (1999), Soil quality – Vocabulary – Part 2: Terms and definitions related to sampling

[18] - ISO VIM (1993), International vocabulary of basic and general terms in metrology.

[19] - NMKL procedure no. 12 (2002), Guide on sampling for analyses of food.

[20] - Thompson, M., & Howarth, R. J. (1976), Duplicate Analysis in Geochemical Practice, *Analyst* 101: 690-698.

[21] - Duncan, A. J. (1974), Quality Control in Industrial Statistics, Richard D. Irwin Inc.

[22] - Miller, J. C. & Miller, J.N. (2000), Statistics for Analytical Chemistry, Ellis Horwood PTR Prentice Hall.

[23] - University of Newcastle, Imperial College and University of Sussex (based on program in new ref 24, copy available from http://www.rsc.org/Membership/-Networking/InterestGroups/Analytical/AMC/Software/index.asp)

[24] - Ramsey M. H. (1998), Sampling as a source of measurement uncertainty: Techniques for quantification and comparison with analytical sources, *Journal of Analytical Atomic Spectrometry* 13, 97-104.

[25] - Miljøstyrelsens Referencelaboratorium (2005), Usikkerhed/fejl ved automatisk prøvetagning af spildevand. Litteraturundersøgelse og forsøgsskitse, Rapport til Miljøstyrelsen.

[26] - Ramsey, M.H. & Ellison, S.L.R. (2007), Measurement uncertainty arising from sampling: a guide to methods and approaches. Eurachem/Eurolab/CITAC/Nord-test/AMC, to be published.

[27] - ISO 3534-1 (1993), Statistics – Vocabulary and symbols.

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

[29] - Thompson, M, & Ramsey, M.H. (1995), Quality concepts and practices applied to sampling – an exploratory study, *Analyst* 120, 261 – 27.

[30] - IUPAC (1990), Nomenclature for sampling in analytical chemistry. Prepared for publication by Horwitz, W., *Pure and Applied Chemistry* 62, 1193 – 1208.

[31] - ISO 78-2 (1999), Chemistry – Layout for standards – Part 2: Methods of chemical analysis.

[32] ISO GUM (1993), Guide to the expression of uncertainty in measurement, 1st Edition, International Organization for Standardisation.

[33] ISO 7002 (1986), Agricultural food products – Layout for a standard method of sampling from a lot.

[34] - 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.

Appendices

Appendix 1 - Groundwater

Measurand				Uncertainty estimation		
Analyte & technique	**Unit**	**Sector & matrix**	**Sampling target**	**Purpose**	**Design**	**Statistics**
Dissolved iron, ICP-AES[3]	mg/L	Environment groundwater	The groundwater near one selected monitoring well in a groundwater body	Measurement uncertainty	Replicate with double split	Relative range

1 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.

2 Scenario and sampling target

A groundwater body which is an important drinking water resource for the city of Århus, the second largest city of Denmark, has through surveillance monitoring been identified as at risk for deterioration of the quality due to intensive drinking water abstraction. An operational monitoring 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[4] and aquitards[5]. The groundwater body as identified is 2 km x 2 km x 10 m, starting 20-30 m below the surface. The natural quality of the groundwater is anaerobic without nitrate, with sulphate and reduced iron, but without hydrogen sulphide and methane. One of the threats to the groundwater body is oxygen intrusion into the aquifer as the result of the water abstraction and concomitant groundwater table draw down.

In the groundwater body, 9 wells had been sampled for chemical analysis during surveillance monitoring, and 6 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.

---

[3] ICP-AES: inductively coupled plasma ionization with atomic emission spectroscopy detection and quantification

[4] Aquifer: underground layer of water-bearing permeable rock, or permeable mixtures of unconsolidated materials

[5] Aquitard: geological formation of layers comprised either of clay or on non-porous rock that restricts water flow from one aquifer to another

Meeting the monitoring objective requires a measurement uncertainty including both sampling and analysis of not more than 10%[6] (quality requirement) corresponding to an expanded measurement uncertainty of 20%. 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 upon 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 was summarised based upon previous monitoring activities (surveillance monitoring). A summary for the two key parameters is shown in Table A1:1. The relative standard deviation here includes variability in time and space as well as measurement (sampling and analytical) uncertainty, i.e. it is the total variation including between target variability.

Table A1:1 Key chemical parameters for 9 wells to the groundwater body, from surveillance monitoring

	Redox potential	**Dissolved iron**
	mV	**mg/L**
Mean	-123	1.11
Relative standard deviation	27%	56%
Main cause of uncertainty	Oxygen impact during sampling and on-line measurement	Filtering of sample prior to analysis

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

3 Sampling procedure

Sampling was done according to the Århus 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 6 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 measurements.

4 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.

[6] The quality requirement is based upon comparison of two means each for two samples, 95% confidence interval, two sided test.

4.1 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 uncertainty in order to identify points of improvement, if required. The validation programme was set up with sampling of 6 wells, two independent samplings per well and 2 sub-samples per sample analysed, see Figure A1:1.

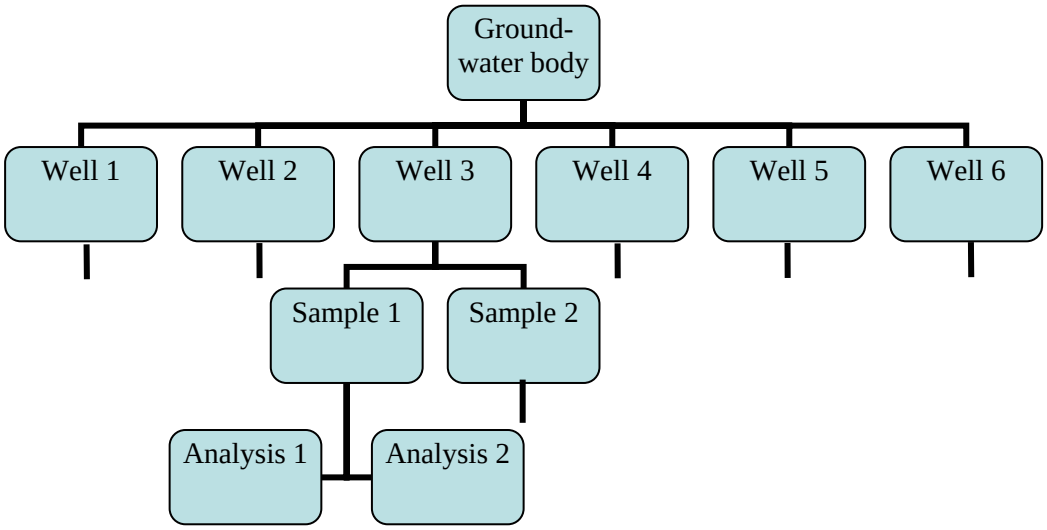


Figure A1:1 Design outline for validation

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

4.2 Quality control

The objective of the quality control programme for the operational monitoring was to ensure that measurement uncertainty did not increase over time during the monitoring. The quality control programme was set up after careful evaluation of the results from the validation study. Quality control was designed to include duplicate sampling and each with duplicate analysis on one of the two annual sampling occasions of the monitoring programme, see Figure A1:2. The quality control programme included 6 sampling occasions in one monitoring well according to the design shown in Figure A1:2.

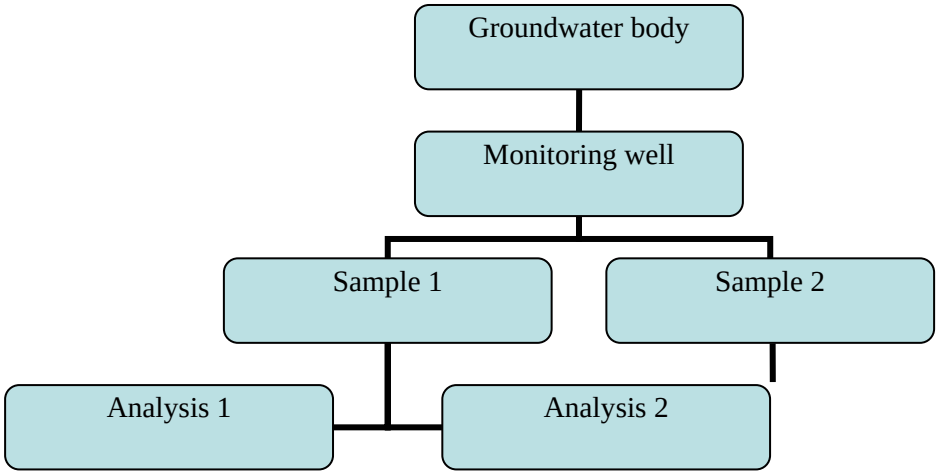


Figure A1:2 Design outline for quality control, shown for one sampling occasion

The sample preparation and analytical set up for the two key parameters (redox potential and dissolved iron) are shown in Table A1:2.

Table A1:2 Preparation and analytical programme

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

5 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, pre-treatment, such as filtering, and transportation. An estimate of the analytical uncertainty alone could be obtained from the laboratory quality control data, see Section 5.2.

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

5.1 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 WTW sensor used for redox potential is described in Table A1:3.

Table A1:3 On-line sensor used for redox potential measurements

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

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

5.2 Laboratory analysis

Analyses were performed at an independent laboratory using accredited methods (ISO 17025) subject to the required quality assurance and analytical quality control. Methods and performance data from analytical quality control are shown in Table A1:4.

Table A1:4 Methods and performance data from quality control for laboratory analyses

	Method	Within series repeatability	Between series reproducibility	Reproducibility	Expanded uncertainty	Detection limit
Iron	ICP-AES	0.95%	4.2%	4.3%	8.6%	0.01 mg/L

The certified reference material (CRM) VKI Metal LL2, nominal 0.200 mg Fe/L was used for quality control with 101.9% recovery (mean for 92 control results).

The replicate data were treated using the relative range method, see Section 6.2. For comparison, uncertainty estimates were calculated by analysis of variances (ANOVA), see Section 6.3, and *robust ANOVA* using ROBAN version 1.0.1 [23, 24].

The applied calculations methods are demonstrated in Table A1:8 of this example. The relative range calculations are easily done using standard spreadsheets, and an example can be downloaded from http://www.samplersguide.com.

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 pre-treatment error. Similarly, redox potential and oxygen contents were checked to correspond in order to control systematic errors.

6 Results

The data set from the validation study is shown in Table A1:8 for dissolved iron with the relative range calculations. The calculations for redox potential in the validation study and for both dissolved iron and redox potential during quality control were done similarly.

The data from the validation study (6 different wells) using range calculations are shown in Table A1:5.

Table A1:5 Relative expanded uncertainty (%, coverage factor 2) for analysis, sampling and between target variability (between wells) as obtained during validation using relative range calculations

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

For comparison, the statistical estimates are shown in Table A1:6 as obtained using ANOVA and *robust ANOVA*.

Table A1:6 Relative expanded uncertainty (%, coverage factor 2) for analysis, sampling and between target variability (between wells) as obtained for dissolved iron during validation using ANOVA and robust ANOVA calculations

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

The statistical estimates obtained with the range statistics during quality control (6 sampling occasions) are shown in Table A1:7.

Table A1:7 Relative expanded uncertainty (%, coverage factor 2) for analysis, sampling and between target variability (between occasions) as obtained during quality control using relative range calculations

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

No groundwater samples had measurements of dissolved oxygen above 0.1 mg/L, and the low redox potential measured (-110 to -200 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).

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

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

In the quality control scheme of monitoring (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, 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, the analytical uncertainty was comparable in validation and in the subsequent quality control (2.1% and 2.5%, respectively). The sampling uncertainty was lower when sampling just one well at different occasions during quality control (3.6-3.8%) than when sampling different wells at the same time during validation (10-15%). The uncertainty between target (variation from one sampling occasion to the next) during quality control was small for dissolved iron (9.9%), but larger for redox potential (23%).

If a continuous control of sampling uncertainty had been required, the control data could have been plotted in a range control chart, see Section 5.2, in order to obtain an early warning of excessive uncertainty (random errors) for each sampling occasion.

Table A1:8 Results and relative range calculations for the validation study, dissolved iron, basic data in bold

Well number	x_{i11}	x_{i12}	x_{i21}	x_{i22}	$D_{i1} = \lvert x_{i11} - x_{i12} \rvert$	$\bar{x}_{i1} = \frac{x_{i11} + x_{i12}}{2}$	$d_{i1} = \frac{D_{i1}}{\bar{x}_{i1}} * 100$	$D_{i2} = \lvert x_{i21} - x_{i22} \rvert$	$\bar{x}_{i2} = \frac{x_{i21} + x_{i22}}{2}$	$d_{i2} = \frac{D_{i2}}{\bar{x}_{i2}} * 100$	$\bar{x}_i = \frac{\bar{x}_{i1} + \bar{x}_{i2}}{2}$	$d_{\bar{x}_i} = \frac{\lvert \bar{x}_{i1} - \bar{x}_{i2} \rvert}{\bar{x}_i} \cdot 100$
	mg/L	mg/L	mg/L	mg/L	mg/L	mg/L	%	mg/L	mg/L	%	mg/L	%
99.474	0.815	0.834	0.912	0.893	0.019	0.825	2.30	0.019	0.903	2.11	0.864	9.03
99.468	1.80	1.83	1.94	1.93	0.030	1.82	1.65	0.010	1.94	0.517	1.88	6.40
99.469	1.69	1.68	1.79	1.77	0.010	1.69	0.593	0.020	1.78	1.12	1.73	5.48
99.916	2.62	2.61	2.83	2.84	0.010	2.62	0.382	0.010	2.84	0.353	2.73	8.07
99.327	1.66	1.63	1.58	1.59	0.030	1.65	1.82	0.010	1.59	0.631	1.62	3.72
99.371	1.52	1.53	1.47	1.50	0.010	1.53	0.656	0.030	1.49	2.02	1.51	2.66
							$\sum d_{i1} = 7.413$			$\sum d_{i2} = 6.750$	$\sum \bar{x}_i = 10.32$	$\sum d_{\bar{x}_i} = 35.36$
							$n_{i1} = 6$			$n_{i2} = 6$	$n_i = 6$	

Analysis	$\bar{d}_{analysis} = \frac{\sum d_{i1} + \sum d_{i2}}{n_{i1} + n_{i2}}$	$d_{analysis} = \frac{7.413 + 6.750}{6 + 6} = 1.18$	$RSD_{analysis} = \frac{d_{analysis}}{1.128}$	$RSD_{analysis} = \frac{1.18}{1.128} = 1.05$		
Sampling	$\bar{d}_{measurement} = \frac{\sum d_{\bar{x}_i}}{n_i}$	$\bar{d}_{measurement} = \frac{35.36}{6} = 5.89$	$RSD_{measurement} = \frac{\bar{d}_{measurement}}{1.128}$	$RSD_{measurement} = \frac{5.89}{1.128} = 5.22$	$RSD_{sampling} = \sqrt{RSD_{measurement}^2 - \frac{RSD_{analysis}^2}{2}}$	$RSD_{sampling} = \sqrt{5.22^2 - \frac{1.05^2}{2}} = 5.17$
Between target	$\bar{\bar{x}}_{variation} = \frac{\sum \bar{x}_i}{n_i}$	$\bar{\bar{x}}_{variation} = \frac{10.32}{6} = 1.72$	$s_{variation} = s_{\bar{x}_i}$ [8]	$s_{variation} = 0.604$	$RSD_{variation} = \frac{s_{variation}}{\bar{\bar{x}}_{variation}} * 100\%$	$RSD_{variation} = \frac{0.604}{1.72} * 100\% = 35.1\%$
					$RSD_{beteentarget} = \sqrt{RSD_{variation}^2 - \frac{RSD_{measurement}^2}{2}}\%$	$RSD_{betweent\,arg\,et} = \sqrt{35.1^2 - \frac{5.22^2}{2}}\% = 34.9\%$

[7] The sum of relative variances is reduced by a factor ½ on $RSD^2_{analysis}$ due to the mean of duplicate analyses being used.

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

7 Comments

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

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

8 Assessment of fitness for purpose

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

9 Reporting and interpretation

Single measurement data for dissolved iron 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.

10 Summary

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

	Expanded uncertainty			**Target variability**
	Sampling	**Analysis**	**Measurement**	**Between wells**
Validation	10%	2.1%	10%	70%[9]
Quality control	3.6%	2.5%	4.4%	9.9%[10]

Acknowledgments

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

[9] In the validation study, between target variability was between wells

[10] In the quality control, between target variability was between sampling occasions

Appendix 2 - Iron Ore

Measurand				Uncertainty estimation		
Analyte & technique	**Unit**	**Sector & matrix**	**Sampling target**	**Purpose**	**Design**	**Statistics**
Total iron XRF	% Fe dried sample	Mining iron ore	Lot – 24 hours production	Uncertainty of sampling	Empirical replicates	Range

1 Scope

The scope is the determination of the sampling uncertainty of the iron content in the produced highly upgraded iron ore pellets at LKAB following sampling according to the ISO protocol 3082 for iron ores.

2 Scenario and Sampling Target

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

3 Sampling procedure

The sampling shown in Figure A2:1 follows ISO 3082[11] for iron ores. Sampling of the pellets in the pelletizing plant is realized by an automatic sampler from a conveyor belt. Every fourth minutes one sample is taken, roughly 300 kg in an hour. After one hour, the sample is automatically divided by splitting and one part is used for screening analysis and one part for chemical analysis, roughly 30 grams. After 8 hours, all 30 g parts of the samples are mixed together into a 240 g laboratory sample and ground automatically. The three 240 g laboratory samples produced during one production day are transported to the analytical laboratory, mixed, splitted and an analytical 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 several parameters of iron ore pellets such as particle size distribution, metallurgical and mechanical test.

[11] ISO 3082:2000. Iron ores – Sampling and sample preparation procedures.

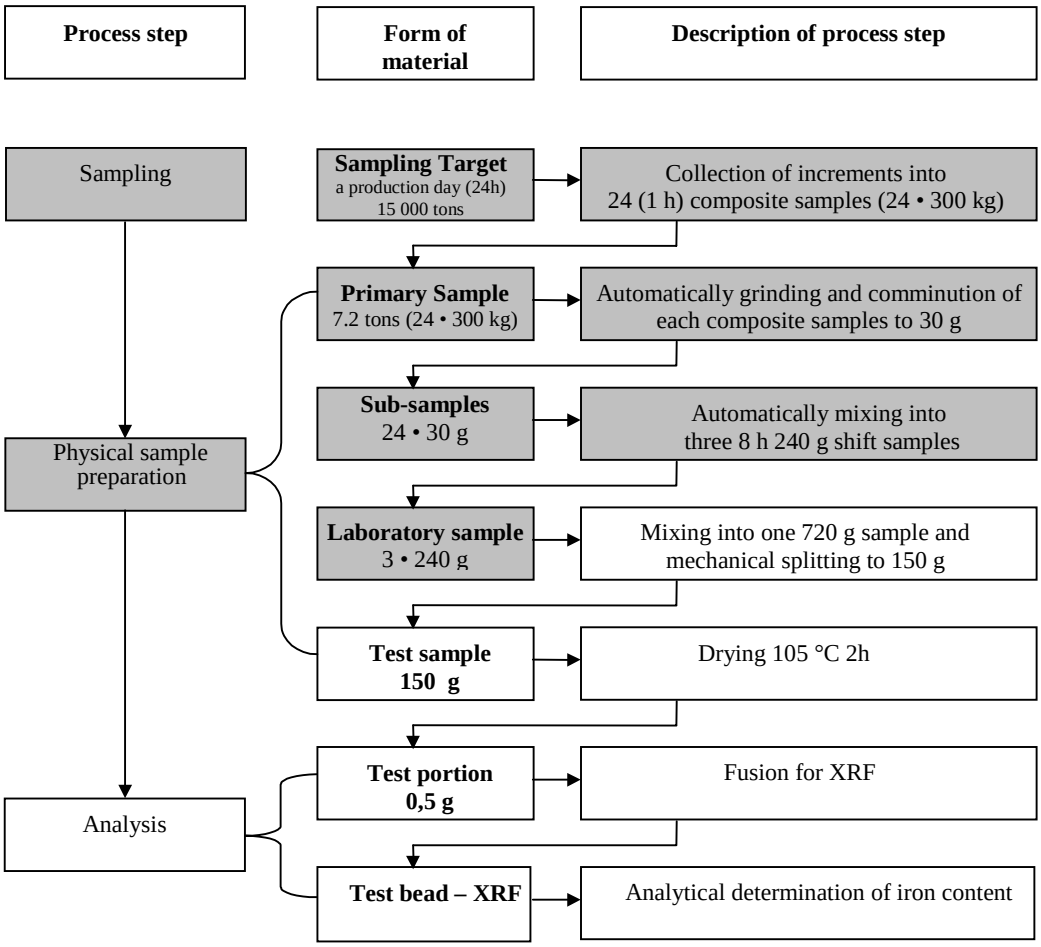


Figure A2:1 Schematic diagram of iron ore sampling and analysis at LKAB, Kiruna

4 Study design – Empirical

The study design follows the main principles of ISO 3085, *Iron ores — Experimental methods for checking the precision of sampling, sample preparation and measurement.* This is an empirical approach with a design using duplicate analyses of the three shift samples - triplicates. However these triplicates are separated in time and if the iron content would vary a lot during a 24 h period this would lead to an overestimate of the sampling uncertainty. This issue is discussed in Section 7 below. This empirical 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.

4.1 Validation

The validation programme was set up using the protocol of ISO 3085 method 1 as a template shown in Figure A2:2.

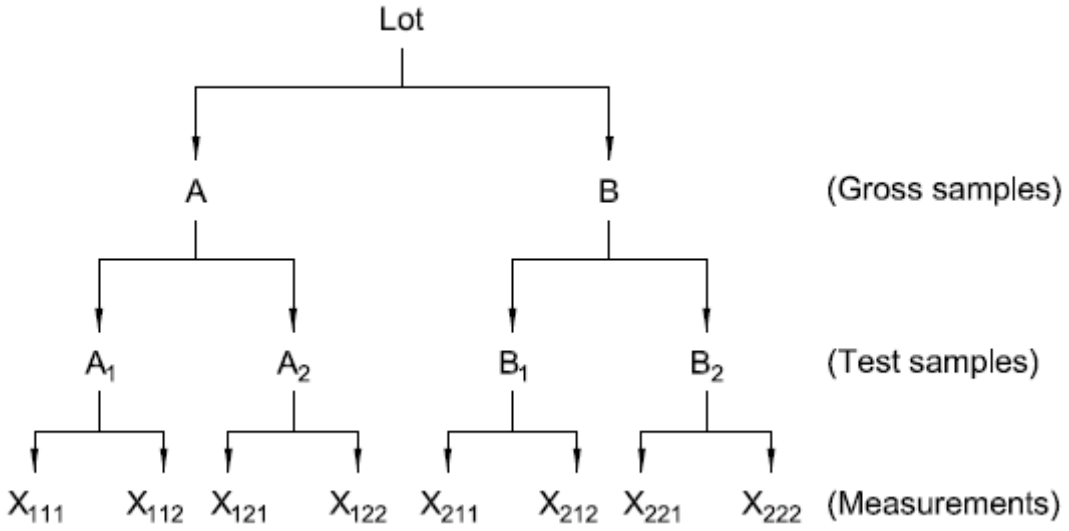


Figure A2:2 Empirical method - experimental design using duplicates – ISO 3085 method 1

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 A2:3.

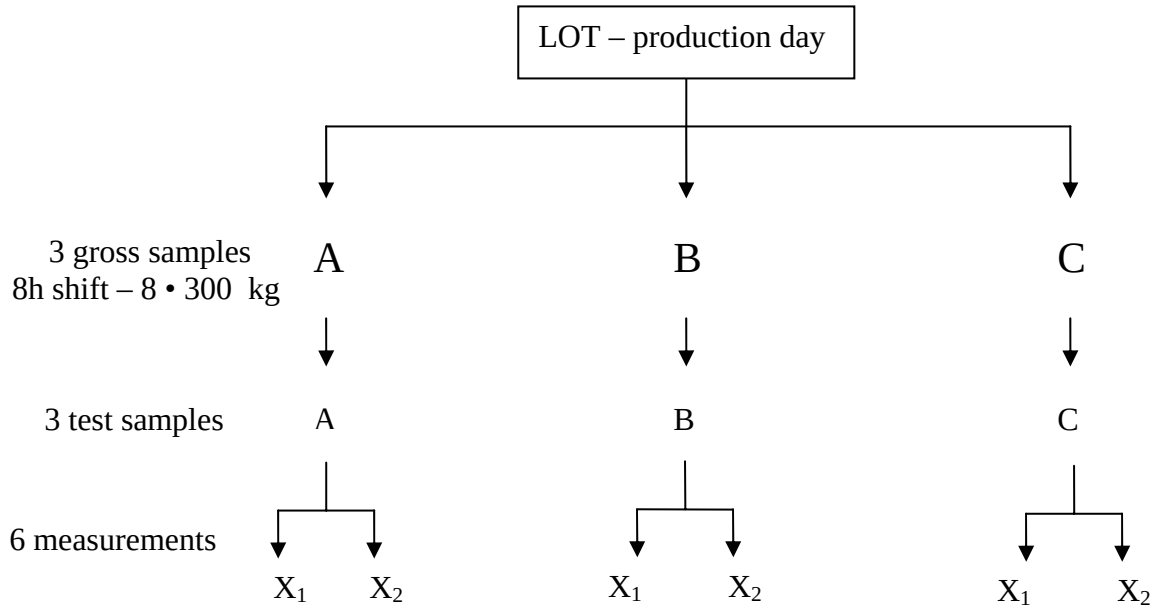


Figure A2:3 Empirical design, modified method 1 of ISO 3085, separating the primary sample into triplicates and after comminution to a test sample, duplicate measurements of each test sample

4.2 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.

5 Sample preparation and analysis

Laboratory sample (the 3 samples are mixed) is split using a mechanical splitter and the test sample is dried (105 °C, 2h). One test portion is analysed with X-Ray Fluorescence (XRF).

5.1 Analyses

XRF – A test portion (0.5g) of the test sample 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 as 100% minus impurities and minus oxygen. The analytical expanded uncertainty reported by the laboratory is 0.20% Fe at a level of 68% iron.

6 Results

6.1 Range calculations and estimation of sampling standard deviation

The overall variation in production including analysis, sampling and product variation over 1 year (September 2004 to November 2005) is 0.16% Fe, expressed as one standard deviation at an iron level of 68% Fe.

This standard deviation consist of the following parts

$s^2_{total} = s^2_{production} + s^2_{sampling} + s^2_{analysis}$ and the measurement part is

$s^2_{measurement} = s^2_{sampling} + s^2_{analysis}$

The results and calculations from data during one week in December 2005 are shown in Table A2:1 to Table A2:3. The number of data is 42 and the raw data is given below.

Table A2:1 Range calculations for the analytical part - iron ore data Appendix 1

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 single analytical measurements.

Table A2:2 Range calculations (duplicate measurement on three separate 8 h shifts during a production day) for measurement (sampling + analytical) part - iron ore data Appendix 1.

Parameter	% Fe	Comment
Measurement – mean range from –triplicates	0.050	
Measurement - standard dev. estimated from range	0.030	s = range/1.693

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.

Table A2:3 Calculations of the sampling part – iron ore data

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

The sampling part of the variation, < 0.01% Fe, is obtained using the following equation – see Section 6.3, split range statistics $s_{sampling} = \sqrt{s^2_{measurement} - \left(s_{analysis}/\sqrt{2}\right)^2}$

6.2 Validation of analysis bias

Comparisons with a long-time study from proficiency testing show no significant analytical bias.

6.3 Validation of sampling and analytical bias

Long-time studies comparing similar sampling with customer sampling the same lot show no significant bias (data obtained from LKAB, personal communication).

6.4 Measurement uncertainty

The repeatability part of the expanded uncertainty (95% confidence interval, k=2) 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 ($\sqrt{0{,}20^2 + 0{,}02^2}$).

7 Comments

For comparison, the calculations were also performed using ANOVA on four samples (with no missing data) from 2005-12-14 – 2005-12-17 with similar results. The $s_{analysis}$ from ANOVA is 0.039% Fe and from range statistics 0.041% Fe and the sampling uncertainty is not significantly different from zero when using an F-test.

The estimated sampling uncertainty is here low. The drawback of estimating based on triplicates separated in time is that it could result in an overestimate due to production variations. An overestimate is not the case here since the sampling uncertainty is estimated to be non significant.

8 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.

9 Reporting and interpretation

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

10 Summary

All values are given expressed as measurement uncertainty at a confidence interval of 95% of the iron concentration for the sampling target of one calendar day. The random part of the expanded analytical uncertainty is 0.08% and the random part of sampling uncertainty is <0.02%. 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%.

Expanded Uncertainty			Target variability
Sampling	**Analytical[1]**	**Measurement**	**Typical production variation**
< 0.02% Fe	0.20% Fe	0.20% Fe	0.32% Fe measured under within lab reproducibility conditions

[1]Obtained from the analytical laboratory at LKAB. Estimated according to Nordtest technical report 537

Acknowledgement
The author is grateful for all assistance and data given by LKAB, Kiruna, Sweden.

Raw data for iron ore

Table A2:4 Sampling one week in December 2005 according to procedure 1- a lot split into three test samples and duplicate measurements of each test 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	68,03	68,02

Appendix 3 - 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 in powder	Food baby porridge-powder	Produced batch	Total measurement uncertainty	Empirical duplicate method	One-way ANOVA

1 Scope

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

2 Scenario and sampling target

In the production of baby (infant) porridge, the vitamin A (retinol) is added as a premix (together with vitamin D and vitamin C). The premix is a minor ingredient. All ingredients are mixed thoroughly before distribution into packages. Earlier analysis indicated a bigger variation in analytical result between packages than expected. A measurement uncertainty of 20 - 30% would be considered acceptable. The question was raised 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 e.g. 3-5 g[12]. One possible explanation of the heterogeneity is that the vitamin premix aggregates in small hot-spots, due to electrostatic interactions with the fruit particles in the porridge powder. The producers recommend a test portion size of 40 – 50 g whenever analysing vitamin A, D and C in baby porridge powder.

Table A3:1 Product data provided by the producer. Data for estimating the "true value" of vitamin A in baby porridge are provided by the producer (Nestlé) of the product chosen for the validation.

Product	**Oatmeal porridge with bananas and apricots (Nestlé)**
Weight of batch, including premix (1 batch = 2 mixing containers)	1092 kg
Weight of added vitamin-premix in batch	1.228 kg
Vitamin A in premix (data from the Certificate of Analysis)	9016 IU/g = 2705 µg/g (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	349 µg/100 g (retinol)
Vitamin A declared as	Retinol - (Sum of trans- and cis-Retinol)

In order to compare the measured vitamin A concentration against declared values and European regulatory thresholds, an estimation of measurement uncertainty is desirable. To determine the random component of the measurement uncertainty, an empirical approach using the Duplicate Method (see Section 9.4.2) is chosen. To estimate the systematic

[12] EN-12823-1 "Foodstuffs – determination of vitamin A by HPLC" indicates a test sample of approximately 2 to 10 g

component a comparison with a reference values is made. The reference value is given in Table A3:1.

3 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 air tight and light protecting bag. It is therefore assumed the degradation of the vitamin A is negligible during normal self 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.

4 Study design – Empirical approach

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

4.1 Validation

Samples are collected on line (just after the filling operation of packages) at random time. Two samples (2 packages, each of approximately 400 g) are collected from each of 10 production units (batches) of one type of baby porridge powder.

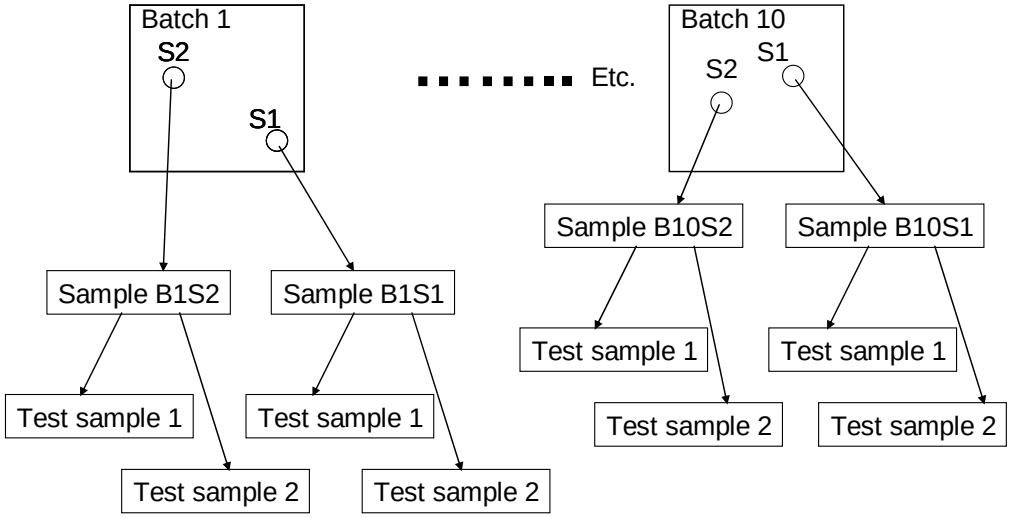


Figure A3:1 Sampling for validation. Two samples are taken from each of 10 production units/batches of the same type of sample.

4.2 Quality control

For quality control (QC) two samples are collected from one batch of each of 8 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.

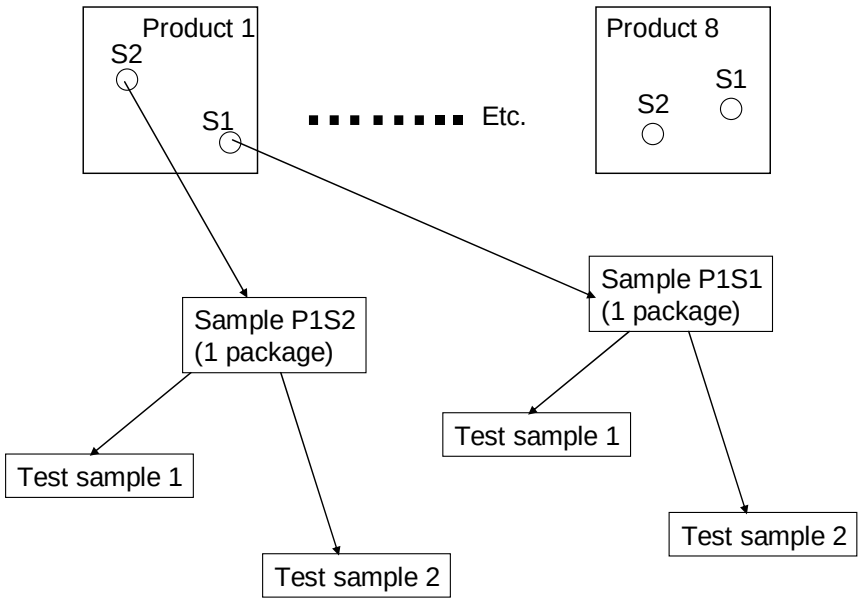


Figure A3:2 Sampling for QC. Two samples are taken from one batch of each of 8 different types of baby porridge.

5 Sample treatment and analysis

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

The laboratory participates in Laboratory Proficiency Tests (FAPAS and Bipea) with good results (in the period 2000 – 2005, |Z-score|<1). The method is validated using a CRM. Data concerning the laboratory performance is given in Table A3:2 below.

Table A3:2 Method and performance data from quality control of vitamin A determined as retinol - laboratory analyses.

Method	EN-12823-1 (HPLC – normal phase column - UV-detection)
Repeatability	2RSD (%) = 6
Within-reproducibility	2RSD(%) = 8
Measurement uncertainty	14% (95% confidence interval)
Recovery	Standard addition, in lab: 90 – 110% Based on laboratory PTs (in period 1999 – 2005), different matrixes: 88 – 113%, mean recovery 100,5%
Limit of Quantification (LOQ)	0.14 mg/kg
CRM used	NIST 2383 – baby food (mixed food composite)
CRM – certified level	0.80 ±0.15 mg/kg (95% confidence interval)
CRM – analysed value	0.77 ±0.14 mg/kg (n=28, 95% confidence interval)

5.1 Secondary sampling

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

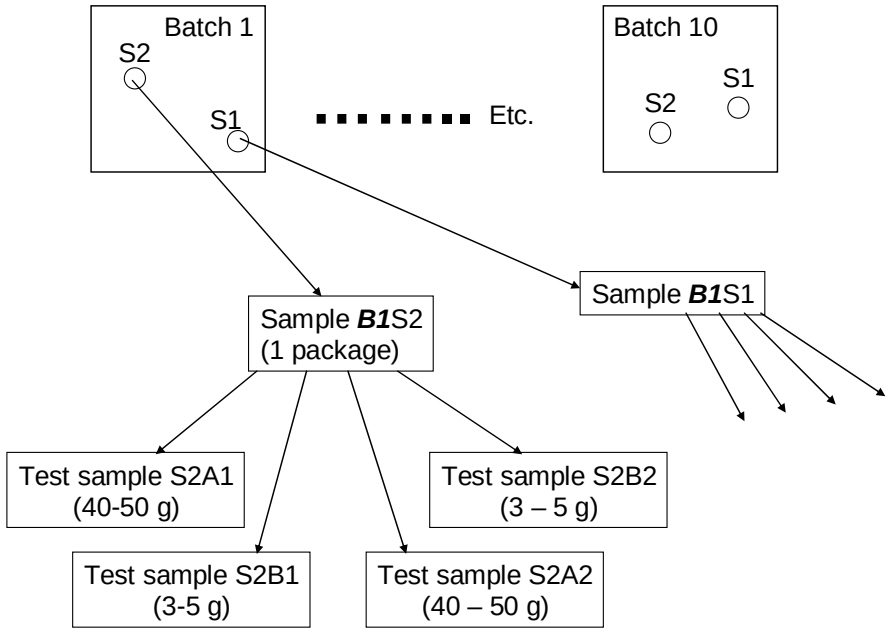


Figure A3:3 Splitting of the primary sample to make 4 test samples.

5.2 Analysis

The analytical method is based on EN 12823-1 (Foodstuffs – Determination of vitamin A by HPLC – Part 1: Measurement of all-trans-retinol and 13-cis-retinol). Retinol is saponified by using ethanolic potassium hydroxide containing antioxidants. Vitamin A is extracted by using

hexane. Analysis is performed by using High Performance Liquid Chromatography (HPLC), with 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).

6 Results

Test sample 40 g – baby porridge

Table A3:3 Validation data - from the 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

S1 and S2: Primary samples from sampling location 1 and 2 of one production batch
A1 and A2: Analyses of duplicate test samples of a primary sample S

Analysed mean value (test sample 40 g): $\overline{X} = 348 \mu g / 100 g$

Test sample 4 g – baby porridge

Table A3:4 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

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

Analysed mean value (test sample 4 g): $\overline{X} = 341 \mu g / 100 g$

6.1 Calculations

The ANOVA calculation can be done by using available tools in Excel, Minitab, SPSS etc. In this study the calculations are done in an Excel spreadsheet and the details of the ANOVA calculations are shown in Table A3:15 and Table A3:16.

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

Table A3:5 Results from ANOVA calculations – uncertainty of analyses - sum of squares of differences, within groups ($SS_{E\text{-}anal}$). For details see Table A3:15.

$SS_{E\text{-}Anal}$ (µg/100g)^2	Degrees of freedom (df_a)	Variance, V_{anal} (µg/100g)^2	Standard deviation, s_{anal} (µg/100g)	Relative standard deviation RSD_{anal} (%)
16595	20	829.75	*28.805*	8.28

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

Table A3:6 Results from ANOVA calculations – uncertainty of sampling - sum of squares of differences SS_{samp}. For details see Table A3:16.

SS_{Samp} (µg/100g)^2	Degrees of freedom (df_{samp})	Variance,V_{Samp} (µg/100g)^2	Standard deviation, s_{samp} (µg/100g)	Relative standard deviation RSD_{samp} (%)
14231	10	296.7	17.22	4.95

Calculation of measurement uncertainty – 40 g test sample

The *RSD* value from the ANOVA calculation can be used as an estimate of the standard uncertainty *u* (%). The analytical laboratory has estimated the analytical standard uncertainty to be 7%, which is lower than the random analytical component for this sample type, 8.28%. The higher value of these two is used in the calculations. Combining the *RSD* values from Table A3:5 and Table A3:6 with Equation 2, $s^2_{measurement} = s^2_{sampling} + s^2_{analysis}$ the results can be written as in Table A3:7.

Table A3:7 Measurement, sampling and analytical uncertainty – 40 g test sample.

	Sampling	**Analytical**	**Measurement**
Uncertainty *u* (%)	4.95	8.28	9.7
Expanded uncertainty *U* (%) = 2**u* With a coverage factor of 2 (i.e. 95% confidence)	9.9	16.6	19

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

The same calculations are used as for test sample size of 40 g (see Table A3:15 and Table A3:16)

Table A3:8 Results from ANOVA calculations – uncertainty of analyses, 4 g test sample - sum of squares of differences, within groups ($SS_{E\text{-}anal}$).

$SS_{E\text{-}anal}$ (µg/100g)^2	Degrees of freedom (df_{anal})	Variance, V_{anal} (µg/100g)^2	Standard deviation, s_{anal} (µg/100g)	Relative standard deviation RSD_{anal} (%)
312206.5	20	15610.325	124.9413	36.68

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

Table A3:9 Results from ANOVA calculations – uncertainty of sampling, 4 g test sample - sum of squares of differences SS_{samp}.

SS_{samp} (µg/100g)^2	Degrees of freedom (df_{samp})	Variance, V_{samp} (µg/100g)^2	Standard deviation, s_{samp} (µg/100g)	Relative standard deviation RSD_{samp}(%)
102860.25	10	-2662.15	$\sqrt{-2662.15}$ Set to zero	Conventionally set to zero

The same calculations are used as for test sample size of 40 g (see Table A3:16) The negative value of $V_{sampling}$ indicates that $s_{sampling}$ is small compared to the calculated value of s_{anal}. In this case, the estimates of $s_{analysis}$ and $s_{sampling}$ using robust ANOVA confirmed the smaller sampling standard deviation; the robust ANOVA estimates were: $u_{sampling}$(%)=6.9% and $u_{analysis}$(%)=30%. As the sampling is identical for the experiments with 40 g and 4 g test samples (and the uncertainty of sampling therefore should be the same), a $RSD_{sampling}$(%) = 5% (≈ 4.95 see Table A3:7) is used as an estimate.

Calculation of measurement uncertainty – 4 g test sample

Using the calculated *RSD*(%) value in Table A3:8 and Table A3:9 as an estimate of the measurement uncertainty and combining with Equation 2, the results can be written as follows:

Table A3:10 Measurement, sampling and analytical uncertainty – 4 g test sample.

	***[i]Sampling**	**Analytical**	**Measurement**
Uncertainty *u* (%) = *RSD*(%)	4.95	36.7	37
Expanded uncertainty *U* (%) = 2**u*	9.9	73.4	74

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

6.2 Effect of the size of test sample on measurement uncertainty

The baby porridge powder looks homogeneous, and therefore a low measurement uncertainty (*u*) is expected. However analyses of the powder indicated in fact a surprisingly large *u* when using a test sample size of 4 g (the CEN-standard EN 12823-1 and other commonly used methods often indicate a test sample size of approximately 2 – 10 g). The producers recommend using a test sample size of 40 – 50 g.

The validation tests gave the following results.

Table A3:11 Comparing measurement uncertainty when analysing test samples of 40 g and 4 g.

Test sample size	**Measurement uncertainty (u_{meas})**	**Expanded measurements uncertainty U_{meas}**
40 g test sample	9.7%	19%
4 g test sample	37%	74%

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

6.3 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 5.2. In the case of baby porridge (40 g test sample) the following calculations can be made:

Warning limit: $WL = 2.83 * \sqrt{(4.95^2 + 8.28^2)}\% = 27\%$

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

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

Table A3:12 Quality control data (µg/100 g)- 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

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 form each laboratory sample.

Table A3:13 Quality control: Calculation of differences D_{ik} and relative difference d_{ik}(%) between samples from a batch where (i) is the number of the batches/product analysed, (j) is the number of samples from each batch and (k) is the test samples analysed of each sample.

Product	Analyses	Sample S1 X_{i1k}	Sample S2 X_{i2k}	$D = \lvert x_{i1k} - x_{i2k} \rvert$	$\bar{x}_{ik}$	$d_{ik}(\%) = (D_{ik} / \bar{x}_{ik}) * 100\%$
P1	A1	322	350	28	336	8
P2		332	358	26	345	8
P3		443	461	18	452	4
P4		318	390	72	354	20
P5		252	265	13	259	5
P6		274	233	41	254	16
P7		206	198	8	202	4
P8		392	375	17	384	4
P1	A2	319	375	56	347	16
P2		317	393	76	355	21
P3		430	388	42	409	10
P4		383	334	49	359	14
P5		219	227	8	223	4
P6		239	217	22	228	10
P7		225	195	30	210	14
P8		335	416	81	376	22

The *d* (%) can be compared directly with the action limit, or is presented in a control chart, see Figure A3:4

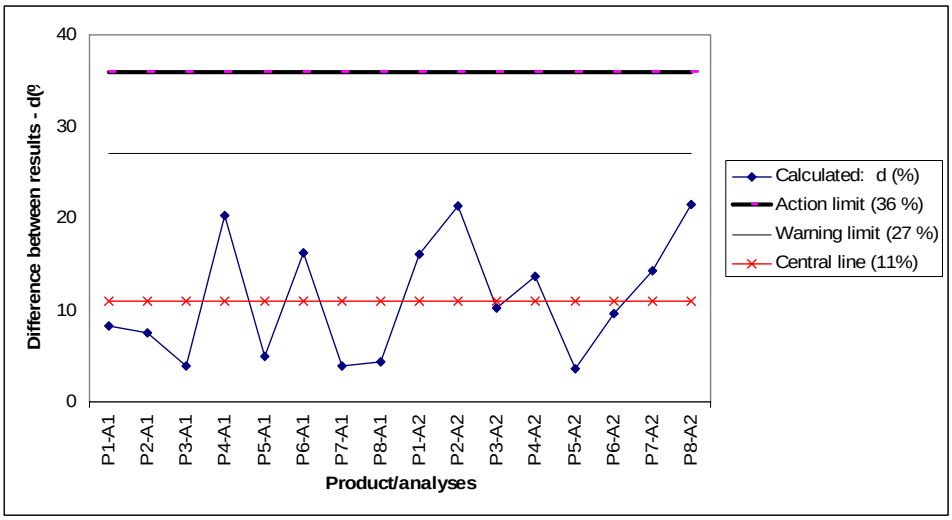


Figure A3:4 Control chart, QC analyses of vitamin A in baby porridge containing cereals and fruits

The control chart in Figure A3:4 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.

6.4 Measurement uncertainty

Sampling uncertainty

Calculations from the validation study gave an expanded sampling uncertainty U_{samp} (%) = 9.9% (40 g test sample – see Table A3:7). The calculated uncertainty does not include contributions to the uncertainty due to "between procedure" and "between samplers" differences.

Analytical uncertainty

Calculation from the validation study gave an expanded measurement uncertainty of analyses ($U_{analysis}$) of 17% for the 40 g test sample. The laboratory reports their own estimation of the analytical uncertainty (see Table A3:2): $2*RSD_{inlab}$(%) = 14%. $2*RSD_{inlab}$(%) is used as an estimate of $U_{analysis}$ in the laboratory. The $U_{analysis}$ found in the validation study was at the same level but still a little bigger than the $U_{analysis}$ reported by the laboratory.

Calculations from the validation study gave an expanded measurement uncertainty $U_{measurement}$(%) = 19% ~ 20% (40 g test sample – see Table A3:7).

6.5 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 A3:14 Certified and analysed data (retinol) for CRM 2383.

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

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

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

In the validation study, the mean value of retinol was determined to be 348 µg/100 g (when using a test sample of 40 g). According to data provided by the producer (see Table A3:1), the "true value" for retinol was calculated to be 349 µg/100 g porridge powder. This gives a recovery of 99.7% 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.*

7 Comments

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

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

8 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.

9 Reporting and interpretation

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

10 Summary

Expanded Uncertainty			Target variability
Sampling	**Analytical**	**Measurement**	**Typical between batch variation** [Calculated as 2*RSD(%) of the mean values of analyses of the batches in the validation study]
9.9%	16.6%	19%	16

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.

Details of the ANOVA calculation, vitamin A in baby porridge

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

ANOVA calculations – uncertainty of analyses - sum of squares of differences, within groups

Table A3:15 Calculation example demonstrating the use of ANOVA for calculating standard deviation of analysis from duplicate samples and duplicate analyses. 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:

S1A1	**S1A2**	**S2A1**	**S2A2**	**S1** $\bar{x}_{i1} = \frac{x_{i11} + x_{i12}}{2}$	**S2** $\bar{x}_{i2} = \frac{x_{i21} + x_{i22}}{2}$	**S1** [13] $\lvert x_{i11} - \bar{x}_{i1} \rvert = \lvert x_{i12} - \bar{x}_{i1} \rvert = D_{i1(\bar{x})}$	**S2** [13] $\lvert x_{i21} - \bar{x}_{i2} \rvert = \lvert x_{i22} - \bar{x}_{i2} \rvert = D_{i2(\bar{x})}$
x_{i11}	x_{i12}	x_{i21}	x_{i22}	$\bar{x}_{i1}$	$\bar{x}_{i2}$	$2 * D_{i1(\bar{x})}^2$	$2 * D_{i2(\bar{x})}^2$
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
$\bar{\bar{X}} = 347.9$				$SS_{E\text{-}analysis} = 2 * \sum_{i=1}^{10} [D_{i1(\bar{x})}^2 + D_{i2(\bar{x})}^2] = 16595$			
[14]) $df_{analysis}$ = (10*2*2-10*2)= 20				$V_{analysis} = SS_{E\text{-}analysis}/df_{analysis} = 16595/20 = 829.75$			
$s_{analysis} = \sqrt{V_{analysis}} = \sqrt{829.75} = 28.8$				$RSD_{analysis} = \frac{s_{analysis}}{\bar{\bar{X}}} * 100\% = \frac{28.8}{347.9} * 100\% = 8.28\%$			

[13] The mean value $\bar{x}_{ij}$ is based on two measurements, therefore the differences, from the mean value to each measurement for the samples, are equal.

[14] The degrees of freedom of analyses $df_{analysis}$ is calculated from $df_{analysis} = i \cdot j \cdot k - i \cdot j$ where (*i*) is number of batches analysed, (*j*) number of samples from each batch and (*k*) number of test samples analysed of each sample.

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

Table A3:16 ANOVA calculations 40 g test sample – uncertainty of sampling - sum of squares of differences

S1A1	**S1A2**	**S2A1**	**S2A2**	**S1** $\bar{x}_{i1} = \frac{x_{i11} + x_{i12}}{2}$	**S2** $\bar{x}_{i2} = \frac{x_{i21} + x_{i22}}{2}$	$\bar{x}_i = \frac{\bar{x}_{i1} + \bar{x}_{i2}}{2}$	**Se footnote [15]** $(D_{i(\bar{x})})^2 = (\bar{x}_i - \bar{x}_{i1})^2 = (\bar{x}_i - \bar{x}_{i2})^2$
x_{i11}	x_{i12}	x_{i21}	x_{i22}	$\bar{x}_{i1}$	$\bar{x}_{i2}$	$\bar{x}_i$	$(D_{i(\bar{x})})^2$
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
$\bar{X} = 347.9$				$SS_{sampling} = 4 * \sum_{i=1}^{10} \left[\left(D_{i(\bar{x})} \right)^2 \right] = 14231$ See further Section 6.3		$SS_{E\text{-}analysis} = 16595$ $df_{analysis} = 20$	
[16]) $df_{sampling} = (10*2-10) = 10$				$V_{sampling} = (SS_{sampling}/df_{sampling} - SS_{analysis}/df_{analysis})/2 = (14231/10-16595/20)/2 = 296.675$			
$s_{sampling} = \sqrt{V_{sampling}} = 17.244$				$RSD_{sampling} = \frac{s_{sampling}}{\bar{X}} * 100\% = \frac{17.224}{347.9} * 100\% = 4.95\%$			

[15] Taking into consideration that the mean value of the batch $\bar{X}_i$ is calculated from two values, the differences from the mean value of the batch to the mean values for each sample are equal.

[16] The degrees of freedom of sampling $df_{sampling}$ is calculated from: $df_{sampling} = i*j-i$ where (i) is number of batches analysed, (j) number of samples from each batch.

Appendix 4 - Conductivity in industrial wastewater

Measurand				Uncertainty estimation		
Analyte & technique	**Unit**	**Sector & matrix**	**Sampling target**	**Purpose**	**Design**	**Statistics**
Electrical conductivity	mS/m	Industrial wastewater	Wastewater outlet spot samples	Measurement uncertainty	Time-series	Variographic analysis

1 Scope

In this example the data will be evaluated using so-called variographic analysis. The scope is to estimate the measurement uncertainty as well as individual uncertainty contributions from inherent heterogeneity, the automatic sampling, pre-treatment and analyses of wastewater in relation to the measurement of electrical conductivity.

It is the intension to focus on the uncertainty contribution from sampling using specific wastewater sampling equipment.

2 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 supervision 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 at determining the concentration or load of pollutants in the wastewater, generally during an extended period of time, for example to monitor compliance with a control limit, to determine trends, to provide data on unit process efficiency or 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 thus in the interest of both the wastewater producer and the supervising authority to assure a uniform and representative sampling and uniform and reproducible results with a known and acceptable uncertainty.

3 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.

4 Study design – Empirical approach

In this example the 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 pre-selected 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 total measurement uncertainty, here called $s_{measurement}$, is the sum of the sampling uncertainty and the analytical uncertainty, see Equation 2:

$$s^2_{measurement} = s^2_{sampling} + s^2_{analysis}$$

Thus, if we can estimate the measurement uncertainty ($s_{meaurements}$) and the analytical uncertainty ($s_{analytical}$), we will be able to estimate by calculation the part of the uncertainty that arises from the sampling process. The estimation of the uncertainty from pre-treatment and analyses of the wastewater samples ($s_{analytical}$) 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 the main text in Chapter 6.4. 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 $s_{measurement}$ at the origin.

For Quality Control (QC), i.e. to calculate the $s_{analytical}$, 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. All measurements of electric conductivity were subjected to ordinary internal quality control by parallel analyses of synthetic quality control samples.

5 Sample preparation and analysis

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

5.1 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 transportable 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 as such (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 sampling sequence the samples were transported to the laboratory. The transportation time was about 45 minutes.

5.2 Analysis

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

6 Results

The resulting data was collected and plotted in time series and as variograms (Figure A4:1 to Figure A4:6). 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, $s_{measurement}$), 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 A4:2 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 A4:1 and as a variogram in Figure A4:2. For detailed information on how to construct a variogram, see Chapter 6.4 in the main text of this handbook.

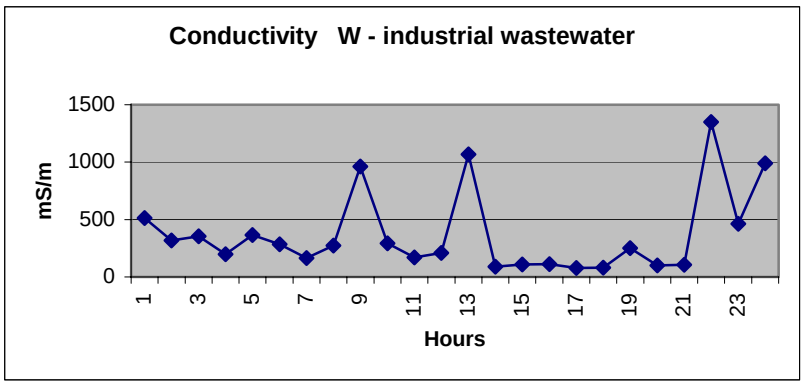


Figure A4:1 Time series W (1 hour increments) – Conductivity.

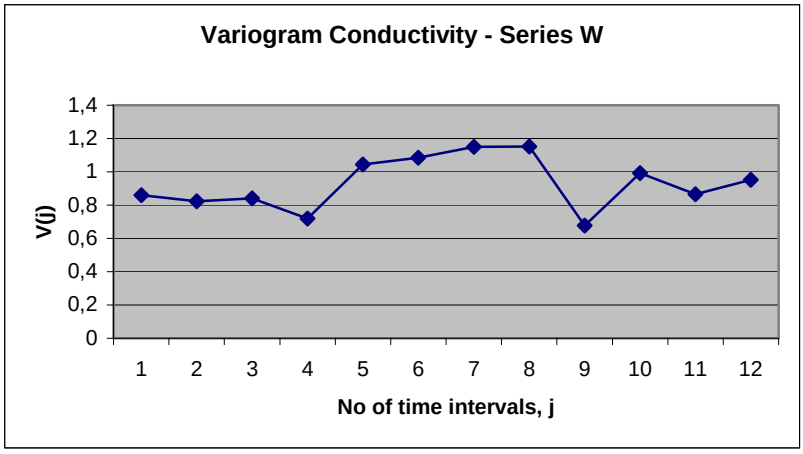


Figure A4:2 Variogram of time series W (1 hour increments).

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 A4:3 and Figure A4:4 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 A4:5 and Figure A4:6. The variograms of series X and Y (Figure A4:4 and Figure A4:6), showing the first 12 points only, makes it possible to estimate the smallest possible sampling error, *V(0)* representing $s_{measurement}$, 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 estimation 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.

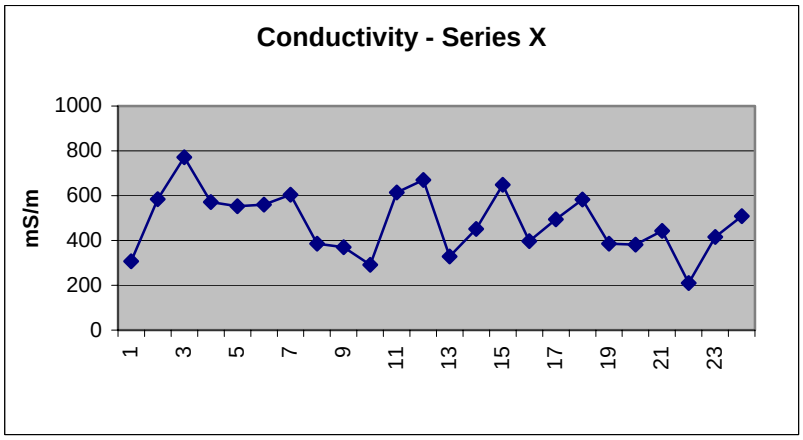


Figure A4:3 Time series X (2.5 minutes increments)– Conductivity.

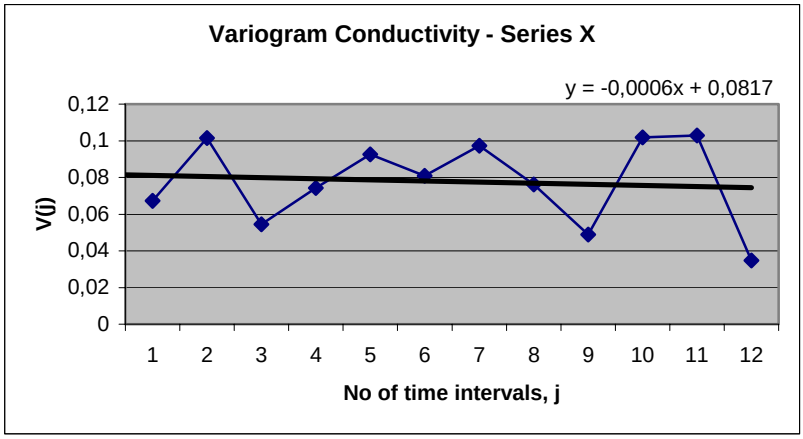


Figure A4:4 Variogram of time series X (2.5 minutes increments).

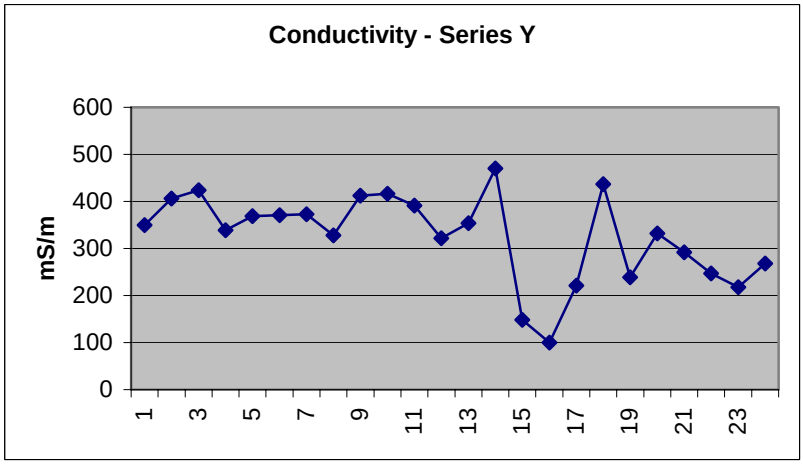


Figure A4:5 Time series Y (2.5 minutes increments) – Conductivity.

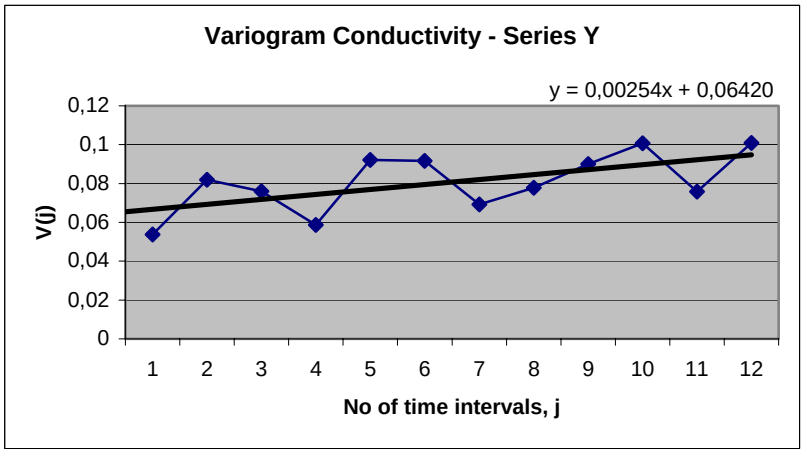


Figure A4:6 Variogram of time series Y (2.5 minutes increments).

Using the information obtained from the fitted lines in the variograms, *V(0)* can be recalculated to the estimate *s(0)* (or $s_{measurement}$) according to Equation 30 and the relative standard deviation, RSD (Equation 31).

The results of the 3 experiments and the above calculations are summarised in Table A4:1.

Table A4:1 Summary of the experimental results of measurement of electrical conductivity from industrial wastewater outlet

	Average mS/m	**$s_{overall}$ mS/m**	**Extrapolated *V(0)***	**Extrapolated *s(0) = $s_{meas.}$***	**Extrapolated *RSD_{meas}***
24 hours (W)	371	356		-	-
60 minutes (X)	481	138	0.0817	138	29%
60 minutes (Y)	326	93	0.0642	82,5	25%
Sample and laboratory variation	340	RSD = 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, see Equation 2 (page 9). Sampling uncertainty is calculated according to Equation 3: $s_{sampling} = \sqrt{s^2_{measurement} - s^2_{analysis}}$

V(0) is a representation of $s_{measurement}$, 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 $s_{measurement}$, from experiments X and Y. In the calculations the higher number from series X ($s_{measurement}$ = 138) is used, in order not to underestimate the uncertainty. If we add information about the analytical uncertainty, in this case a relative standard deviation of 1.3% taken from repeated analysis of the 10 L samples, we can estimate the uncertainty from the sampling:

$s_{analytical} = 0.013*481 = 6.07$

$s_{sampling} = \sqrt{138^2 - 6.07^2} = 138$, corresponding to 29% of the average 481 mS/m.

Sampling thus contributes to virtually all of 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, factors like representativity of the sample are not included (i.e. does the sample taken truly represent the average concentration of the whole wastewater stream in the period under investigation?).

7 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.

8 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 uncertainty of analysis is negligible in this context. The high uncertainty probably reflects that the design of the sampling site is not optimal. The study can therefore be used to identify and quantify the effect of sub-optimal design of the sampling site.

9 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 given as e.g. 100 mS/m ± 29 mS/m.

10 Summary

The measurements in the outgoing industrial wastewater shows 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 particle load might also be an important source of error.

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**	
58%	2.6%	58%	-

Table A4:2 Results of measurement of electric conductivity (EC) from sampling and analyses. Sampling of 24 samples over 24 hours (W) and sampling of 24 samples over 60 minutes (2.5 minutes increments) (X and Y) by the use of the automatic water sampler.

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
Average	371		481	326
s	356		138	93
RSD	96%		29%	29%

norden
Nordic Innovation Centre

Return address:
Nordic Innovation Centre,
Stensberggata 25
NO-0170 Oslo, Norway

NORDTEST

NORDTEST is a Nordic Innovation Centre brand offering competence and expertice in the field of harmonizing of norms and methods,a large Nordic net-work of experts, more than 650 recommended Nordic testing methods and 550 published technical reports.

www.nordicinnovation.net

Nordic Innovation Centre

The Nordic Innovation Centre initiates and finances activities that enhance innovation collaboration and develop and maintain a smoothly functioning market in the Nordic region.

The Centre works primarily with small and medium-sized companies (SMEs) in the Nordic countries. Other important partners are those most closely involved with innovation and market surveillance, such as industrial organisations and interest groups, research institutions and public authorities.

The Nordic Innovation Centre is an institution under the Nordic Council of Ministers. Its secretariat is in Oslo.

For more information: **www.nordicinnovation.net**