

Markku Ohenoja

MMEA WP2: Measurement and sampling uncertainty – a literature review



CLEEN LTD
ETELÄRANTA 10
P.O. BOX 10
FI-00130 HELSINKI
FINLAND
www.cleen.fi

ISBN 978-952-5947-73-1
ISSN XXXX-XXXX

Cleen Ltd.
Research Report D2.1.13

Markku Ohenoja
Control Engineering, Faculty of Technology, University of Oulu

MMEA WP2: Measurement and sampling uncertainty – a literature review



mmea

Measurement, Monitoring and Environmental Assessment

Cleen Ltd
Helsinki 2015

Report Title: Measurement and sampling uncertainty – a literature review

Key words: Uncertainty, estimation, modeling approach, empirical approach, GUM, analytical, sampling, TOS, Variogram, Monte Carlo Simulation, Fuzzy, Random-fuzzy, Dynamic measurements, Virtual measurements, Digital signal processing, Quality control

Abstract

All measurements include some level of uncertainty. The uncertainty should always be taken into account when using a measurement result in decision making or process control. There are well defined methods of evaluating the measurement uncertainty both at instrument level and at system level. Metrology is a branch of science that focuses on the uncertainty estimation at instrument level and many guides regarding the topic have been published. Partly the same statistical methods are also applied to empirical system level uncertainty evaluation and some practical guides are well known and widely used for analytical measurements. Theory of sampling extends the uncertainty evaluation and error avoidance to the sampling process as well. This review presents the central features of the methods related to measurement uncertainty and sampling error estimation. Also the uncertainties in digital signal processing and virtual measurements, and the alternative methods in evaluating those, are addressed.

Helsinki, May 2015

Table of contents

Table of contents	2
1. Introduction	3
1.1. Basic notation.....	4
1.2. The measurement target	4
1.3. Sample types	6
2. The analysis of measurement uncertainty	7
2.1. Modeling approach.....	9
Six steps of evaluation of the uncertainty	9
Uncertainty arising from sampling	12
2.2. Empirical approach	13
2.3. Alternative methods fulfilling the GUM	16
Monte Carlo method	18
Fuzzy approach	19
3. Theory of Sampling.....	21
3.1. Sources of sampling and estimation error	21
3.2. Sampling strategies.....	24
3.3. Estimating the sampling error	26
3.4. Variographic analysis.....	27
Variogram	27
Calculation of the variogram.....	28
Variograms in practice	31
4. Applications.....	35
4.1. TOS and variographic analysis	35
4.2. Uncertainty estimation for virtual instruments.....	38
5. Fit-for-purpose.....	41
6. References	44

1. Introduction

Uncertainty, as a parameter associated with the result of a measurement, characterizes the dispersion of the values that could reasonably be attributed to the measurand. All measurements include some level of uncertainty. The measurement uncertainty depends on the unknown factors and the inaccurate knowledge of the known factors. Hence the measurement result is actually an estimate of the measurement target. Therefore also the uncertainty should always be mentioned with the analytical result. Even greater amount of uncertainty is originated from the sampling process. In pure and applied science uncertainty may be used to judge the consistency between experiment and theory, different measurements, and different theories (Cox, Harris & Siebert 2003).

Metrology is a branch of science investigating the measurement uncertainty. Traditionally a metrological interpretation of a measurement process focuses on instrument level of a measurement. Thus a correct sampling is assumed and the focus is more on a fundamental representation of the uncertainty related to the analysis process. Theory of sampling (TOS) is a scientific theory originally imposed by Pierre Gy over fifty years ago. TOS focuses on the determination of the uncertainty components related to sampling process and aims to point out the proper sampling procedures and frequencies i.e. ensuring the representativeness of the samples. Therefore TOS is a system level representation of measurement uncertainty, or error. The principles of both theories are presented in the following two sections. Section 4 gives some practical examples of utilizing the tool presented, especially in process and environmental applications and intelligent measurements. Finally, Section 5 presents a very important aspect of quality control, fit-for-purpose (FFP) evaluation of a measurement.

There is a recent literature review on measurement uncertainty published in IEEE Transactions on Instrumentation and Measurement (da Silva Hack, ten Caten 2012). The review is however restricted into selected journals between 2004 and 2010 and collects the approaches and calculation methods from 114 scientific articles. This review and tutorial include also journals outside the ones in (da Silva Hack, ten Caten 2012) and extends the view into the earlier development of measurement uncertainty estimation. The material studied in this review especially concerns with the inclusion of sampling process as an

additional source of uncertainty, an aspect that is not considered in the review of (da Silva Hack, ten Caten 2012).

1.1. Basic notation

The theories involve a carefully chosen notation. The basic terms are presented here and some additional definitions are given in more appropriate context. The original material being subject to sampling is called *a lot*. As a result of correct sampling, *a sample* is extracted from the lot. If the sampling process is non-correct, the extracted product is *a specimen*. Another extraction from the lot is *an increment*, which is a partial sample consisting of a single fragment, or a single or several groups of fragments to be used in *a composite sample*. A composite sample consists of combined sub-samples (increments) where only the composite sample is to be analyzed. (In some context, composite samples are called bulk samples (Esbensen, Paoletti & Minkkinen 2012)) *A fragment* is the smallest physically separable particle (molecule, mineral etc.) in the lot. *A group* is a number of spatially correlated or spatially coherent fragment associations, usually denoting the material occupying the extracted volume in the sampling tool. (Petersen, Minkkinen & Esbensen 2005)

Once the sample is extracted from the lot, *a measurement* can be taken in order to determinate the value of a quantity via (automatic) set operations. *A (measurable) quantity* refers to a quantity in general sense (e.g. length, mass, resistance) or to a particular quantity (e.g. length of a given rod) and is an attribute of a phenomenon, body or substance that may be distinguished qualitatively and determined quantitatively. *A value (of a quantity)* is the magnitude of a particular quantity including a unit of measurement and a number. The particular quantity subject to the measurement is called *a measurand*, e.g. vapor pressure of a given sample of water at 20°C. (JCGM100 2008)

1.2. The measurement target

Defining the sampling target: the properties and characteristics of the material e.g. in a certain area or time period are of interest, it is considered as a sampling target. Each sampling target will generate one reported measurement result with its uncertainty. A composite sample is collected from a number of primary samples. The single measurement value is obtained from this composite sample. The value of uncertainty is affected by the number of increments (primary samples) that are taken. This is in contrast with the case

of several distinct primary samples, each of them measured separately and then using the mean value as a measurand. The standard error related to that is actually the standard error of the mean value, not the uncertainty estimate for the measurement. The standard error of the mean value can be decreased by taking more distinct primary samples. (Ramsey, Ellison 2007)

Defining the sampling target can also be misunderstood. The final measurement result can easily be judged as a measurand describing of the analyte in the batch of material, whereas an analysis may call it as a measurand of the laboratory sample. Another example is “contaminant concentration at a factory outlet at the time of sampling” vs. “the average contaminant concentration over a year”. The first viewpoint includes the effects of sampling while the other does not. (Ramsey, Ellison 2007)

The sampling can be taken from different types of objects, or lot types. The most important distinction is between the 0-D lot and lots of higher dimensionality. 0-D lot has no internal correlation between the individual increments i.e. the increments are completely random. Basically, assuming a 0-D lot, means that the whole object is to be measured. 1-, 2- or 3-D lots the individual increments are fixed and spatially correlated along the defining dimension(s), either in space or time. Process streams are usually considered as 1-D lots as they are moving or stationary stream of particulate material (e.g. powders in a conveyor belt), a moving or stationary string of fluids (e.g. fluids in pipelines), or a moving or stationary stream made of discrete chronological units (bags, truck loads etc. from a production line). (Petersen, Esbensen 2005) Figure 1 depicts the difference between lots of different dimensionalities. The colored increments are the extracted ones.

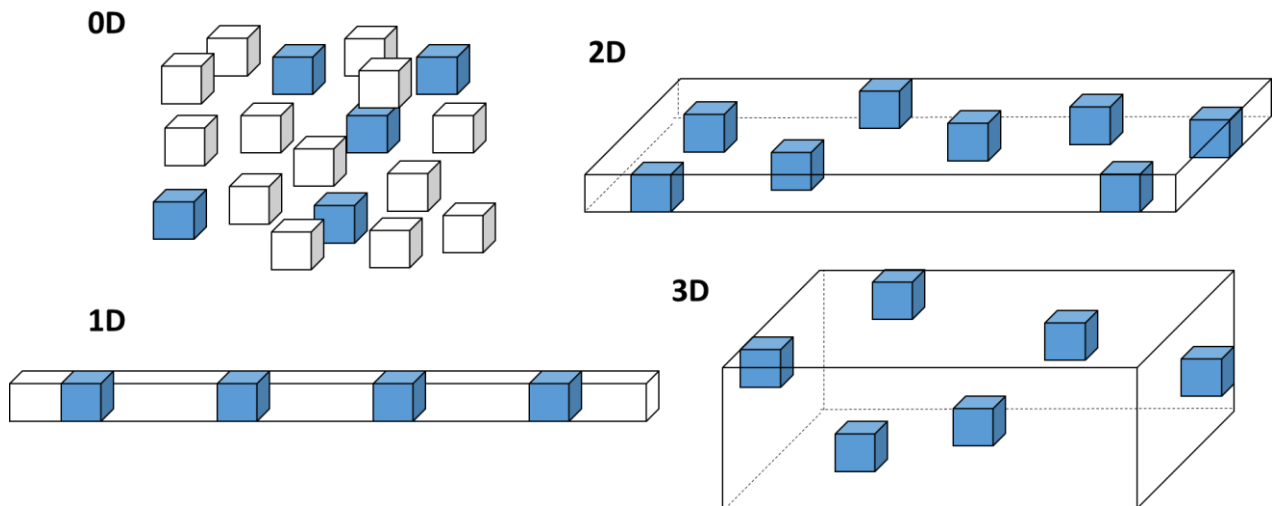


Figure 1. Illustration of lot dimensionality, redrawn from (Wagner, Esbensen 2011).

1.3. Sample types

Composite samples are mixed from several single samples in order to give a more reliable sample of the average, or general properties of the measurement target. Single samples only give transient properties. Commonly composite samples are used for routine monitoring purposes and considered without taking into account the measurement uncertainty. Even the interpretation of an expert involves a risk of an unreliable decision as the uncertainties cannot be quantified (Lambkin, Nortcliff & White 2004).

2. The analysis of measurement uncertainty

Joint Committee for Guides in Metrology and ISO Technical Advisory Group on Metrology have published a guide to the expression of uncertainty in measurement (JCGM100 2008), also referred as GUM. The purpose of the guide is to promote full information on how uncertainty statements are attained and to promote a basis for the international comparison of measurement results used within standardization, calibration, laboratory accreditation and metrology services. The ideal method for evaluating and expressing the measurement uncertainty should be universal i.e. applicable to all kind of measurements and to all types of input data used in measurements. (JCGM100 2008) Another guidance, the EURACHEM/CITAC guide (Ellison, Williams 2012) presents the uncertainty estimation, following the GUM principles, in analytical measurements. The approach taken in GUM concerns measurements with a well-defined physical quantity (measurand) that can be characterized by an essentially unique value. However, if the phenomenon can only be represented as a distribution of values or is e.g. time-dependent, the measurands are the set of quantities describing that distribution or that dependence.

An analysis of a measurement should cover the measurement target (gradients, homogeneity, disturbance due to the measurements, temporal variations), the measurement method (reproducibility, variations affecting the target and equipment, representativeness, errors in sampling), the measurement device (calibration, reference standard, drift, resolution, sensitivity, non-linearity, interactions), the measurement conditions and the measurer. An uncertainty analysis may point out the weak points of the measurement and give information on factors which may help to improve the quality of the measurement. Uncertainties in analytical processes are listed in (Ellison, Williams 2012). GUM gives a following list of the possible sources of uncertainty in a measurement:

- Incomplete definition of the measurand
- Imperfect realization of the definition of the measurand
- Nonrepresentative sampling — the sample measured may not represent the defined measurand
- Inadequate knowledge of the effects of environmental conditions on the measurement or imperfect measurement of environmental conditions
- Personal bias in reading analogue instruments

- Finite instrument resolution or discrimination threshold
- Inexact values of measurement standards and reference materials
- Inexact values of constants and other parameters obtained from external sources and used in the data reduction algorithm
- Approximations and assumptions incorporated in the measurement method and procedure
- Variations in repeated observations of the measurand under apparently identical conditions.

The uncertainty estimate can be estimated empirically (empirical, experimental, retrospective, top-down approach) by repeating the whole measurement procedure or mathematically (modeling, theoretical, predictive, bottom-up approach) by quantifying the sources of uncertainty individually and constructing a model. Applicability of the two approaches depends on the sampled system. Also a dual approach can be used and many practical solutions involve both of these elements. (Ramsey, Ellison 2007) The empirical approach includes all sources of uncertainty, but cannot identify the sources. However, it is applicable when there is no prior knowledge of the nature of the material. If prior knowledge is available i.e. the application is well-characterized, the modeling approach may be more cost-efficient as it does not require extensive experimental studies. Modeling results as a long-term solution whereas empirical approach is more valuable when testing different sampling targets. (Ramsey, Ellison 2007)

The analysis of a measurement uncertainty (GUM) is intended for correct application of measurement and sampling procedure i.e. small errors. However, the sampling can also produce gross operator error (e.g. misuse of the measurement protocol, highly heterogeneous material) (Ramsey, Ellison 2007). Sampling is often a major source of uncertainty in test results and usually the uncertainty arising from taking the field samples is much greater than errors associated with preparation, handling and analytical and data analysis (Lambkin, Nortcliff & White 2004). Traditionally, the sampling and analysis processes are treated independently in uncertainty estimation. More recently, the sampling has been considered in the same context with analysis and a guide (Ramsey, Ellison 2007) has been published. The uncertainty can only be estimated if there is an understanding of both the analytical and the sampling process. The complete understanding can easily be achieved if one

person is responsible for all stages of the measurement process. In many cases, the sampling process and the analytical processes are made in different locations and by different people or organizations. Therefore, all of the parties involved should have a common guidance of the measurement process in order to estimate the uncertainty. (Ramsey, Ellison 2007)

2.1. Modeling approach

Although called a modeling approach, the calculation of the measurement result is based on input data i.e. the measured values of single or repeated observations, a judgment based on experience, corrections, information brought from external sources, such as calibrated measurement standards, certified reference material and reference data from handbooks. (JCGM100 2008) This section introduces the original, well-established methodology for uncertainty estimation with the modeling approach. Some further considerations concerning sampling are also given. Alternative modeling approaches based on numerical simulation and possibility distributions are given in Section 2.3.

Six steps of evaluation of the uncertainty

The evaluation of the uncertainty comprises on six steps: defining a measurement model, estimating the standard uncertainties of the model inputs, calculating the effect (sensitivity) of the uncertainties to the measurand, determining the correlation between inputs, combining the uncertainty components and, finally, calculating the expanded uncertainty. (JCGM100 2008) and (Heinonen 2010) are used as a source material on introducing the following six steps:

Step 1

The measurand is usually determined through a functional relationship rather than measured directly (e.g. analyzer gives an electrical current signal as an output and a calibration function transfer the current signal into the concentration of some component in the gas examined). The measurement model is an equation (or a set of equations) including all input quantities x_i affecting significantly the estimate y and/or the combined standard uncertainty $u_c(y)$. It is notable that magnitude of a correction can be zero but it still has an uncertainty and therefore needed to be included in the model. Typical input quantities are for example the signal itself, calibration equation, drift, resolution

and ambient conditions. The equation describing the measurement Y consisting of input quantities X_i can be written as:

$$Y = f(X_1, X_2, \dots, X_N)$$

The measurement result, which is an estimate of the measurand Y , or the output estimate, denoted as y , is obtained from the input estimates x_i

$$y = f(x_1, x_2, \dots, x_N)$$

Step 2

After all the affecting inputs are identified, their standard uncertainties need to be determined. Standard uncertainties and variances are required in order to compare the inputs. The evaluation of standard uncertainty of an input can be done by the statistical analysis of series of observations (Type A evaluation of uncertainty) or by other means than the statistical analysis using a priori distributions (Type B evaluation of uncertainty).

Type A evaluation uses the arithmetic mean of the quantity (\bar{q}) as an estimate and is applicable when the number of measurements is high enough and normal distribution can be assumed. If so, the experimental standard deviation $s^2(q)$ and the estimate for the variance of the mean $s^2(\bar{q})$ can be calculated:

$$s^2(q) = \frac{1}{n-1} \sum_{i=1}^n (q_i - \bar{q})^2$$

$$s^2(\bar{q}) = \frac{s^2(q)}{n}$$

The standard uncertainty of quantity q , denoted as $u(q)$, equals the square root of the estimate for the variance of the mean i.e. $s(\bar{q})$. {Note: the notation for type A quantities is q but for type B quantities x .}

Type B evaluation is typically applied for:

- uncertainties of values and drifts of reference standards
- uncertainties of environmental quantities
- uncertainties from specifications of instrument
- uncertainties from literature values

- uncertainty due to the method or the calculation
- uncertainty due to staff
- uncertainties from calibration certificates

These quantities require a priori information about their range and distribution. If limiting values are known and all the values have an equal probability, rectangular distribution can be used. Sometimes also triangular distributions or U-shape distributions are applied. Once the distribution and the range of values are chosen, the standard uncertainty $u(x_i)$ can be obtained from the associated equations.

Step 3

Next the contributions of $u(x_i)$ (and $u(q)$) to the uncertainty of y need to be determined. The effect is denoted by the sensitivity coefficient c_i i.e.

$$u_i(y) = |c_i|u(x_i)$$

The sensitivity coefficient can be determined from partial derivatives ($c_i = \partial f / \partial x_i$), by numerical methods or experimentally ($c_i = \Delta y / \Delta x_i$).

Step 4

The calculation of the combined standard uncertainty must consider if the input quantities are independent or if they have mutual dependence. The covariance can either increase or decrease the uncertainty. The covariance of $x_i = F(q_l)$ and $x_j = G(q_l)$ depending on the same quantities q_l is calculated as:

$$u(x_i, x_j) = \sum_l \frac{\delta F}{\delta q_l} \frac{\delta G}{\delta q_l} u^2(q_l)$$

Step 5

The combined standard uncertainty is calculated as a square root of the sum of the squared standard uncertainties multiplied with their sensitivity coefficients. The latter term in the equation is only needed if correlated input quantities are found.

$$u_c(y) = \sqrt{\sum_{i=1}^N \left(\frac{\delta f}{\delta x_i}\right)^2 u^2(x_i) + 2 \sum_{i=1}^{N-1} \sum_{j=1}^N \frac{\delta f}{\delta x_i} \frac{\delta f}{\delta x_j} u(x_i, x_j)}$$

Step 6

Expanded uncertainty U is the standard uncertainty multiplied by a coverage factor k i.e. $U = k u_c(y)$. Typically a coverage factor $k=2$ is used corresponding roughly a 95% level of confidence, if normal distribution is assumed. However, the coverage probability should not be stated, as the traditional approach may not justify that claim (Kacker, Jones 2003).

Although the modeling approach is a long-term solution, the model should be revised if the observed data demonstrate that model might be incomplete. In order to avoid incompleteness in the first place, all relevant quantities should be varied to the fullest practicable extent so that the evaluation of uncertainty can be based as much as possible on observed data. Whenever feasible, the evaluations of uncertainty should include empirical models of the measurement founded on long-term quantitative data, the check standards and control charts that can indicate if a measurement is under statistical control. (JCGM100 2008)

Uncertainty arising from sampling

Instead of analytical measurements only, the modeling approach can also be used for estimating the uncertainty arising from sampling. For the analysis part, the uncertainty can be monitored by using repeated measurements and certified reference materials, but it is difficult to apply these methods to the sampling of materials that are variable and reference materials are not available (Lambkin, Nortcliff & White 2004). Identifying the potential sampling errors leads to a cause-and-effect modeling and fish-bone diagrams. One such example is given in (Lambkin, Nortcliff & White 2004). A cause-effect diagram collecting the possible error sources of the sampling process in that study is presented in Figure 2. Another example is given in (Kurfürst et al. 2004), where the modeling approach based on a reference database was used on the estimation of uncertainties resulting from sampling, sample preparation and analysis of soil samples.

Another model describing the sampling errors comes from the Theory of Sampling. Especially the empirical part of TOS is treated in detail in Section 3. The Gy's formula for determining the sampling uncertainty for particulate systems from the first principles is presented e.g. in (Gy 2004b, Minkinen 2004, JCGM100 2008). This model can predict the uncertainty arising from sampling and can be used as an additional input to the measurement model, but requires a good knowledge of the material (particle) properties.

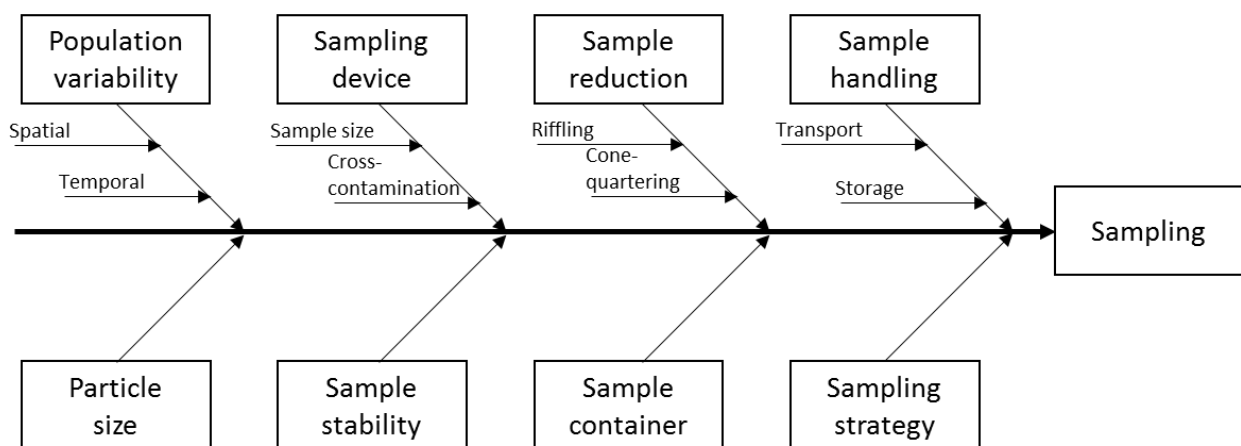


Figure 2. Cause-effect diagram for the sampling phase, redrawn from (Lambkin, Nortcliff & White 2004).

2.2. Empirical approach

The empirical approach relies on overall reproducibility estimates from inter-organizational trials, internal method validations and quality control. As these trials are often expensive and time-consuming, an alternative method involving e.g. duplicate sampling and analysis has been found suitable especially when heterogeneity is the main source of uncertainty. (Lyn et al. 2007) It cannot separate the uncertainty of individual sources, but some general classification into four sources of error can be made. These are the random errors and systematic errors arising both from the sampling processes and analytical processes, which are usually called the sampling precision, analytical precision, sampling bias and analytical bias. Three of them can be estimated with well-established methods, but estimating the sampling bias is more challenging. (Ramsey, Ellison 2007)

Sampling and analytical precision can be estimated e.g. by the duplication method, i.e. repeating some proportion of the samples/analyses. An estimate of the analytical bias can be taken from the validation of the analytical method or estimated by measuring the bias on well-matched certified reference material. Also sampling bias might be estimated using a reference sampling target in some cases. An alternative approach on estimating sampling bias includes inter-organisational sampling trials. If an evidence (e.g. a prior knowledge of the chemical or physical nature of the sampling target, a prior information from earlier measurements on complete batches) that systematic effects are small and under good control can be found, it may be unnecessary to estimate the uncertainty of systematic errors. (Ramsey, Ellison 2007)

The empirical approach uses a statistical model describing the relationship between the measured and true values of the analyte:

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

where x is a single measurement of one sample (composite or single) from one particular sampling target. The term ε_{target} describes the between-target variation and is recommended to be included in the estimation of sampling uncertainty. The measurement variance, assuming that the sources of variations are independent, can be described by

$$\sigma_{meas}^2 = \sigma_{sampling}^2 + \sigma_{analysis}^2$$

The total variance, including the sampling, is given by

$$\sigma_{total}^2 = \sigma_{target}^2 + \sigma_{sampling}^2 + \sigma_{analysis}^2$$

In both equations, the variances σ^2 can be approximated with the statistical estimates of variance, s^2 and the standard uncertainty of the measurement can be calculated by

$$u = s_{meas} = \sqrt{s_{sampling}^2 + s_{analysis}^2}$$

(Ramsey, Ellison 2007) lists five empirical methods applicable to the estimation of uncertainty:

- Duplicate method; described below

- Sampling protocols; bias between protocols can be detected
- Collaborative trial in sampling (CTS); bias between different samplers, when one protocol is applied
- Sampling proficiency test (SPT); bias arising either from the sampling protocol or the sampler
- Variographic analysis; described in detail in Section 3.4.

In the duplicate method, the duplicate samples are obtained using a single sampling protocol and by a single sampler and it can be used for estimating the sampling and analytical precision. The sampling bias need to be estimated separately, or assumed to be negligible. Analytical bias can be obtained, if certified reference materials are included in the analytical process. The duplicate method uses 10%, or at least 8 duplicate samples of the sampling target. In the analysis stage, the test samples are also duplicated. This approach is called as a balanced design and it is illustrated in Figure 3. The analytical results of the duplicate samples are then applied in analysis of variance (ANOVA) or range calculation in order to estimate $s_{sampling}$ and $s_{analysis}$. The random component of the measurement uncertainty (u) can then be calculated according to model described above.

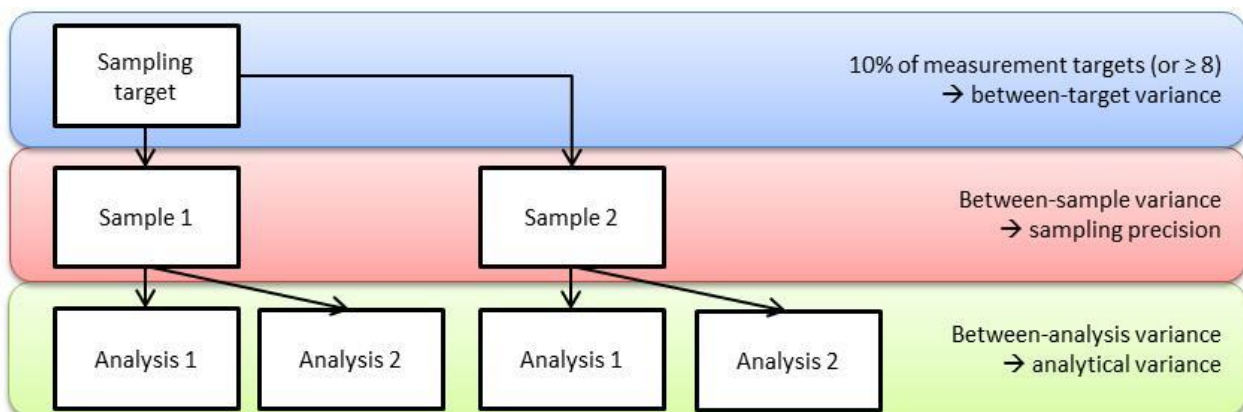


Figure 3. Balanced experimental design for empirical estimation of uncertainty, redrawn from (Ramsey, Ellison 2007).

The duplicate method often gives a reasonable reliable estimate of uncertainty as usually between-operator and between-protocol effect are much smaller than those caused by heterogeneity. A routine application of duplicate sampling can also be used for monitoring the ongoing sampling quality or representativeness. (Ramsey, Ellison 2007)

An illustrative example of the modeling approach and the empirical approach is given in (Lyn et al. 2007). They determined the measurement uncertainty, including both the analysis and sampling, of a contaminant in particulate matter. Due to the nature of the system, the modeling approach followed the Gy's model. The empirical approach used the duplicate method and RANOVA. They found that the empirical approach provided a substantially lower uncertainty estimate than modeling approach (22.5% vs. 136%, respectively). One reason for the deviation is that the empirical method aims to estimate the uncertainty of a case-specific particular sampling target, whereas the modeling approach may use parameters that are not applicable to the present target under investigation. They concluded that empirical approach is applicable to routine estimation, but only the modeling approach can offer tools for reducing too high sampling uncertainty.

(Lyn et al. 2003) extended the empirical approach to include also the error caused by the physical sample preparation step. They also used spike recovery trial data for the estimation of systematic errors in sample preparation. The methodology proposed was primarily aimed for method validation procedures where it may provide useful information about the uncertainty associated to sample preparation.

2.3. Alternative methods fulfilling the GUM

As the importance of uncertainty estimation has become more and more recognized, and the well-established methods have been proposed, the methodology keeps up developing further. GUM has been revised already several times and supplements have been published and are being prepared. The science community have pointed out some weaknesses of the original GUM and also proposed alternative methods, based on alternative theories, to measurement uncertainty estimation. Also new measurement applications, e.g. imaging measurements and virtual instruments have driven the urge for some revision of the uncertainty estimation methodology. This section gives a short overview of these alternative methods, which are also aimed at fulfilling the requirements of GUM.

For example, GUM does not completely follow the classical statistics, but also uses the Bayesian statistics. (Kacker, Jones 2003) proposes a revision of the methods in GUM in order to provide a consistent viewpoint based on Bayesian statistics. Additionally, the original approach in GUM uses the first-order Taylor

expansion (i.e. linear approximation) of the measurement equation for the propagation of the confidence intervals. The confidence interval is represented by the standard deviation of a probability density function and the combination of those results as a Gaussian distribution based on the Central Limit Theorem. These assumptions may not be valid in some cases, e.g. if only few contributing terms exist or some of them are non-Gaussian. (Lovett, Ponci & Monti 2006) Alternative methods to the propagation of distributions have been presented. For example, (Lovett, Ponci & Monti 2006) applied the polynomial chaos theory for the representation and propagation of measurement uncertainty, which enables also describing probability distribution functions differing from the normal distribution. A numerical simulation method, Monte Carlo Method, uses the probability distributions of the input variables instead of their variances in the propagation. The Monte Carlo Method is presented below and has already been accepted by the standardization organizations.

Some of the uncertainty components might not be probabilistic, but systematic or unknown, and should be modeled using alternative models and methods (Reznik, Dabke 2004). These kinds of input quantities are especially seen in intelligent measurements and other DSP (digital signal processing) measurements. They are indirect measurements, where the final measurement result is formed from a set of input signal samples through some processing. In the case of discontinuous relationship between the final measurement and the result of single measurement, or with high number of inputs (thousands in DSP), GUM might be troublesome (Ferrero, Salicone 2003). Hence, approaches based on Fuzzy variables have been proposed to overcome such problems and some of the literature is presented below. Some systematic error component can also be treated with the general methodology by increasing the expanded uncertainty to include bias. (O'Donnell, Hibbert 2005) demonstrates with simulation how the bias could be treated with this kind of approach.

Yet another important type of uncertainty estimation problem concerns dynamic measurements. The calculation of stationary uncertainty can be extended to transient measurements, where the amplitude of the uncertainty of the input variables is time-dependent. The measurement system can be said to be a dynamic system if the response time of the system is comparable to or slower than the changes of the measurand. In a static system, the response of the system is much faster than the changes of the measurand. (Hessling 2009)

Examples of dynamic systems are accelerometers and transducer systems measuring force or pressure. Dynamic metrology is proposed and presented in (Hessling 2009) and references therein. It comprises the linear modeling of the measurement system and estimating the derivatives of the model in transform domain, hence utilizes the methods known in signal processing and control engineering.

Monte Carlo method

The traditional GUM approach presented in Section 2.1 assumes a Gaussian distribution or a scaled and shifted t-distribution of the output quantities (JCGM101 2008). A subclause to GUM, GUM supplement 1 (JCGM101 2008), presents an alternative approach on determining the probability distribution functions using the Monte Carlo method (MCM). The probability distribution function can be based on Bayes' theorem or the principle of maximum entropy. MCM approach does not require the classification of the quantities into Type A or Type B, nor does it require the calculation of the sensitivity coefficients (and partial derivatives). MCM also provides a numerical representation of the distribution function and therefore the evaluation is not restricted onto assumed symmetric distributions as in traditional approach. As the resulting distribution might be asymmetric, a coverage factor may not be centered on the estimate. Hence the choice of the coverage factor requires some additional consideration when using MCM. (JCGM101 2008) The limitations of MCM might be the runtime with poor computational power, the difficult selection of the proper functions of inputs if the data is inaccurate or there is a lack of process understanding, and finally, the dependency between the accuracy of the result and the quality of the random number generator (Herrador, Asuero & González 2005). (JCGM101 2008) lists the uncertainty evaluation problems, where the traditional approach may fail, but MCM approach should be applied:

- the magnitudes of the contributory uncertainties are not approximately on the same level
- the partial derivatives of the model are difficult or inconvenient to calculate
- the probability density function for the output does not follow neither of the assumed distributions
- an estimate and the associated standard uncertainty are approximately on the same magnitude

- the models are arbitrarily complicated
- the probability density functions for the input quantities are asymmetric.

It is also recommended that MCM is applied in order to validate the results of the traditional GUM. MCM can also be used to account the dynamic effects of the uncertainty (Hessling 2009). In addition to (JCGM101 2008), the reader is guided to (Cox, Harris & Siebert 2003, Herrador, Asuero & González 2005) for more information about applying Monte Carlo method in uncertainty estimation.

Fuzzy approach

Fuzzy variables and sets can be used to express the uncertainty of a measurement. Possibility theoretical approach allows considering both the random effects and the systematic effects. Therefore it extends the original GUM that assumes the systematic error is negligible or corrected in the measurement under investigation. Especially the uncertainty estimation of virtual instruments, or DSP based measurements, has driven the need for accounting systematic effects. The uncertainty estimation of such measurements is treated more detailed in Section 4.2.

A fuzzy expression of measurement uncertainty, compatible with GUM, is given in (Mauris, Lasserre & Foulloy 2001). Some examples of the applicability of the fuzzy approach are given in (Reznik, Dabke 2004). (Urbanski, Wąsowski 2003) claim that the original fuzzy expression by (Mauris, Lasserre & Foulloy 2001) describes properly only the propagation of systematic error. Comprehensive description of both the statistical and systematic error component should therefore be based on random-fuzzy variables. (Urbanski, Wąsowski 2003) then adds that there is no theory comprising fuzzy random variables and they propose the t-norm based arithmetics to describe the propagation of statistical and systematic components.

The lack of generality of the abovementioned fuzzy approaches is pointed out by (Ferrero, Salicone 2003, Ferrero, Salicone 2004). They propose a random-fuzzy approach framed within the theory of the evidence as it combines the possibilistic and the probabilistic approaches. The concepts of the belief measure and the plausibility measure are used. In (Ferrero, Gamba & Salicone 2004), an example of measurement uncertainty estimation of a DSP-based instrument is given by the means of both the random-fuzzy variables and GUM. They also suggest that the usage of random-fuzzy variables enables an online

estimation of the uncertainty. A more recent summary of the theory of evidence and uncertainty estimation is given in (Salicone 2013).

3. Theory of Sampling

Theory of Sampling (TOS), developed by Pierre Gy, extends the analysis of sampling uncertainty as it includes both the errors of correct sampling as well as the errors of incorrect sampling. Incorrect sampling errors are due to incorrectly designed sampling equipment or procedures. Correct sampling errors arise from the heterogeneity of the material in sampling targets. Usually, a properly taken increment is called 'a sample' from which one can draw a valid conclusion concerning the properties of the whole lot. 'A specimen' is a sample with incorrect sampling, where such conclusions cannot be made. (Minkkinen, Esbensen 2009) The definition of a correct sampling process or procedure is: "All fragments, or group of fragments, or increments of the lot, must have an equal, non-zero probability of ending up in the sample, while elements foreign to the lot must have a zero probability of ending up in the sample. The increment or the sample must not be altered in any way."

TOS aims to answer the following questions so that a certain level of uncertainty is achieved with minimal costs; 1) How should the samples be cut from the process stream? 2) How many samples should be taken? 3) How often samples should be taken? 4) Which sampling strategy should be used? The basis of the estimation of the uncertainty is the identification of the error sources. (Paakkunainen, Reinikainen & Minkkinen 2007)

The complete description of TOS is given in textbooks by Gy (French and English). A more compact description and the history of TOS can be found from the five articles published in 2004 in Chemometrics and Intelligent Laboratory Systems (Gy 2004a, Gy 2004b, Gy 2004c, Gy 2004d, Gy 2004e). This review and tutorial avoid repeating that material and mainly relies on other references.

3.1. Sources of sampling and estimation error

In TOS, there are seven error sources of sampling, presented in Figure 4. Most methods used in estimation of sampling error deal only with one of them, the Fundamental Sampling Error (FSE). In order to understand the origin and consequences of correct (and incorrect) sampling, an overview of all error sources is needed. (Petersen, Minkkinen & Esbensen 2005)

The total error of an analytical result is the Global Estimation Error (GEE), consisting of two fundamental contributions: the Total Analytical Error (TAE) and the Total Sampling Error (TSE). According to [Petersen et al. 2005] TAE

is should be familiar to all chemists and other analysts. TAE can be estimated with the approach presented in Section 2. TSE arises from material heterogeneity and the sampling process itself, where the former one sets the limits on the achievable performance of the results. Often, the latter one is purposely increased as the emphasis of sampling is on obtaining a small sample with minimum time and effort with least expense and most direct or simplest procedure. This kind of procedure, most probably leading to a non-representative sample, is denoted "Grab sampling". (Petersen, Minkkinen & Esbensen 2005)

All sampling operations include Fundamental Sampling Error (FSE) and Grouping and Segregation Error (GSE). FSE is due to the lot heterogeneity, i.e. is inherent to the material properties and it depends on the number of critical particles in the samples. Especially for solids, powders and particulate materials at low concentrations, FSE can be very large. On the opposite, for homogeneous gases and liquids, FSE is very small. (Minkkinen 2004) Only way to reduce FSE is to physically (e.g. comminution of particles) improve the lot characteristics so that more representative sampling can be obtained. FSE can be estimated only to an order of magnitude and only to a lot that can be thoroughly mixed before sampling. However, it is very useful for any sampling stages after primary sampling. If the required uncertainty level is predefined, FSE can be used on estimating the variance of a given sampling step and minimum sample size. (Petersen, Minkkinen & Esbensen 2005)

GSE is related to both the sampling process (e.g. due to sampling mass) and the material heterogeneity as it arises from the composition and spatial distribution heterogeneity of the lot material. GSE is at its minimum when individual fragments are selected to form the sample. Naturally, in practical situations this is impossible as usually the neighboring fragments of an extracted fragment is more likely to be extracted too etc. However, the contribution of grouping can be minimized by decreasing the size of the increments. The contribution of segregation can be reduced by a proper mixing. Ideal mixing would minimize GSE and therefore give a lower possible residual heterogeneity. If this is not possible, composite sampling is preferable. (Petersen, Minkkinen & Esbensen 2005)

Abovementioned errors are unavoidable i.e. they also occur in correct sampling. Incorrect Sampling Errors (ISE, PME in the figure) consists of

Increment Delimitation Error (IDE), Increment Extraction Error (IEE or IXE) and Increment Preparation Error (IPE). IDE occurs when the actual shape of the extracted increment deviated from the correct geometrically delimited increment, e.g. the entire slice of the whole stream cannot be extracted. Usually, complete sampling is enabled and IDE is avoided by stopping the stream (conveyer belt) for calibration purposes. IEE arises e.g. from particles falling on the wrong side of the sampling device so that some particle sizes or gas components are over-/underrepresented in the extracted sample. IPEs are random events of human error, contamination, adhesion, moisture uptake etc. between the actual extraction of the sample and the analysis and they do not follow any statistical distributions. (Petersen, Minkkinen & Esbensen 2005)

In Figure 4, also abbreviations SWE and PSE exist. SWE, the sample weighting error is caused if the flow-rate of the process stream is ignored. Point selection error (PSE) is raised from using a discrete sampling of a continuous process with random and cyclic drifts. PSE is therefore divided into long-term point selection error (PSE1) and periodic point selection error (PSE2). The size of total PSE depends on the sample selection strategy and the degree of auto-correlation. (Paakkunainen et al. 2007) In 1-D sampling, the point selection errors may also be called the time fluctuation error (TFE) and the cyclic fluctuation error (CFE) (Esbensen et al. 2007).

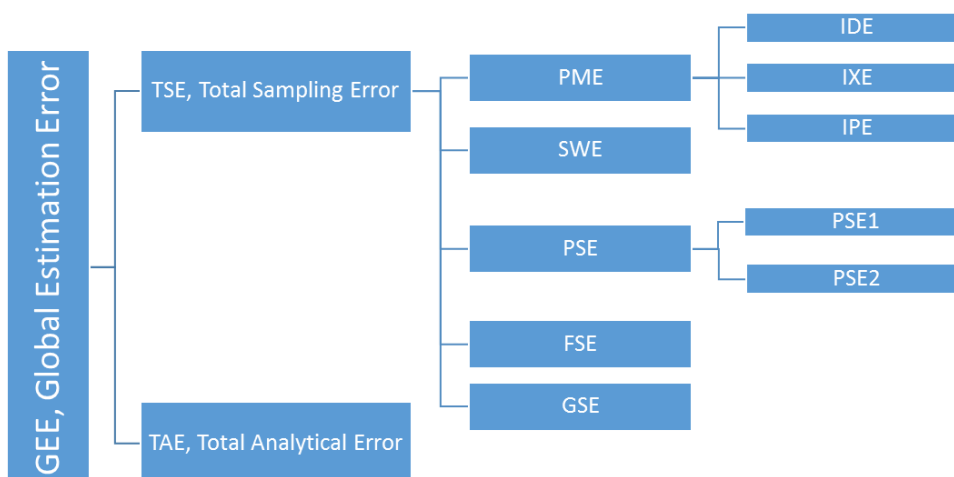


Figure 4. Sampling error components according to Gy, redrawn from (Paakkunainen, Reinikainen & Minkkinen 2007).

3.2. Sampling strategies

There is a guideline of seven practical sampling principles (sampling unit operations) to be followed in order to identify and minimize the errors of sampling (Petersen, Minkkinen & Esbensen 2005). Table 1 presents the principles and their demand.

Table 1. Sampling unit operation according to (Petersen, Minkkinen & Esbensen 2005).

	UNIT OPERATION	DEMAND
1	Heterogeneity characterization	when new sampling operation is initiated
2	Mixing before all further sampling steps	always
3	Prefer composite sampling	always
4	Only use representative mass reduction	always
5	Comminution whenever necessary	if needed
6	Perform variographic analysis of 1-D heterogeneity	when new sampling operation is initiated
7	Turn 2-D and 3-D lots into 1-D equivalents if possible	in order to apply TOS

In addition to above presented principles, a sampling plan should be optimized in order to minimize FSE and GSE. In the case of 1-D process streams, two more sampling errors related to the short- and long-range periodic fluctuations should also be minimized. A useful tool for the latter one is variographic analysis presented in Section 3.4. (Petersen, Minkkinen & Esbensen 2005)

Another list of the crucial aspects of sample collection in order to produce the best practicable effort to obtain a representative sample is given in (Lambkin, Nortcliff & White 2004):

- identification of the population to be sampled;
- ensuring an adequate number of samples;
- determining the frequency and timing of sampling events;
- determining the sampling pattern (regular, random or stratified random); and,
- sub-sampling.

The sampling plan depends on the required accuracy level, the heterogeneity of the material, the degree of mixing, variability between the batches and time. Statistical analysis of the data can be used to determine the aspects concerning the sampling plan (Lambkin, Nortcliff & White 2004):

- the mass and number of increments;
- the mass of a composite sample;
- the intervals for sampling on a mass or time basis;

- whether sampling should be random or fixed-interval; and,
- the bias and the precision associated with routine sampling.

The three most fundamental sampling strategies, or patterns are: regular sampling, stratified sampling and random sampling. In random sampling, the samples are taken randomly from the whole lot. In systematic sampling, the lot is divided into sub-lots of equal size and then a sample is taken at a fixed interval. In stratified sampling, the lot is divided in a similar manner, but samples are taken randomly from each sub-lot. Figure 5 illustrates the sampling strategies.

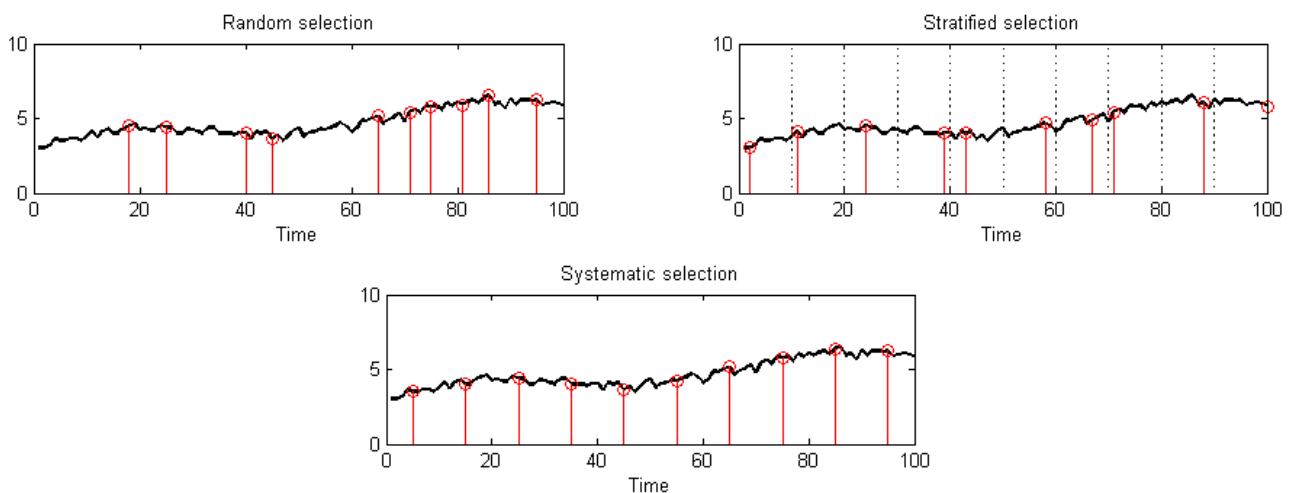


Figure 5. Sample selection strategies.

Random sampling usually gives the highest standard deviation of the lot mean and the systematic sampling the lowest. However, if the sampling interval is equal to the harmonic multiple of the process periodic fluctuations, the systematic sampling will be biased and provide the highest standard deviation. Therefore, stratified sampling is recommended if no reliable information about the periodicity is available. (Paakkunainen, Reinikainen & Minkkinen 2007)

All sampling strategies can also be used to form a composite sample if estimating the mean value of the lot is of interest. The sampling strategy affects also the variance of the mean (based on the composite sample) and the uncertainty can therefore be decreased with a proper selection of the sampling strategy.

3.3. Estimating the sampling error

The global estimation error can be expressed as [Esbensen et al. 2012]:

$$GEE = TAE + PSE1 + PSE2 + \sum_n (FSE_n + GSE_n + IDE_n + IEE_n + IPE_n)$$

In some cases, FSE can be estimated if the necessary material properties are known using e.g. particle distribution models etc. In practical sampling, the conditions of correct sampling can seldom be obeyed and FSE cannot therefore be estimated solely. If the lot consist of a single unit with a randomly distributed property i.e. the sampling target is 0-D, and the correct sampling is used, the variance of FSE can be estimated from the variance of the mean of n replicate samples (Paakkunainen, Reinikainen & Minkkinen 2007). Also for lots consisting of several well-defined sub-lots of equal or unequal sizes, a variance estimate can be calculated from the replicate samples using the analysis of variance (ANOVA). Such lots are for example bags, barrels, wagon loads when there is no serial correlation between these units. (Paakkunainen, Reinikainen & Minkkinen 2007)

If the sampling target is 1-D, and if the sample preparation and size reduction by splitting are carried out correctly, fundamental sampling error models can be used for estimating the variance components generated by these steps. Further, if the expectance value for the number of critical particles in the sample can be estimated as a function of sample size, the uncertainty of the sample can be estimated using Poisson distribution or binomial distribution as a sampling model. (Minkkinen 2004) FSE estimation based on such models is demonstrated in (Korpelainen et al. 2002, Minkkinen 2004, Petersen, Esbensen 2005).

(Wagner, Esbensen 2011) give a comprehensive example of identifying sampling errors and estimating the uncertainty related to sampling procedures. The example comes from the CO₂ emission of a coal fired power plant including the whole pathway from fuel to emission. They found that the power plant had some un-ideal sampling devices and principles both in the primary sampling stage and also in the mass reduction stage. After the identifying and estimation of the sampling errors, the uncertainty of the CO₂ emission was increased i.e. the emissions were underestimated with the default sampling procedures. They concluded, and their results showed, that there is clearly

room for unspecified uncertainties as the EU guidelines do not account representative sampling.

In most sampling cases, such as process sampling, the combined contribution of FSE, GSE, TAE and IPE can be estimated experimentally from the variographic analysis. Together they form the minimum possible error (MPE). The variographic analysis is presented in detail in the following section with references given to wide range of applications. Variographic analysis is particularly useful in situations where large-scale spatial and/or temporal variation can be quantified and modeled (Ramsey, Ellison 2007). The distinction between variographic analysis and time-series analysis is that time-series analysis is primarily concerned with behavior that can be fully modeled by a sum of periodic functions, without specific MPE regards to the errors (Esbensen et al. 2007).

3.4. Variographic analysis

Variogram

Variographic analysis, or statistics of auto-correlated series can especially be used on identifying the point selection errors i.e. long-term and periodic fluctuation errors PSE1 and PSE2. Variogram characterizes the standard deviations of the sampling with different sampling intervals and/or sampling strategy. If no auto-correlation exists, the sampling strategies (random, systematic, stratified) provide the same sampling variance. (Paakkunainen, Reinikainen & Minkkinen 2007) Another important advantage of variographic modeling is that the compositing takes place in spreadsheets and does not require renewed sampling (Minkkinen, Esbensen 2013).

An example of an experimental variogram, the result of a variographic analysis, is presented in Figure 6. The descriptions of the basic characters of the variogram are given before the actual calculation is presented. There are three fundamental features that carry all the important primary information related to sampling errors and process dynamics (Esbensen, Paoletti & Minkkinen 2012):

- The Sill; the sill is the upper-bounding value in y-axis. It represents the variance calculated from all experimental heterogeneity values, for all lags and practically corresponds to the maximum variance of the data series under investigation. (Esbensen et al. 2007) adds that the time series is no longer correlated after the sill.

- The Range; the range is the lag distance beyond which the variogram function levels off and reaches the sill. Samples taken at lags below the range are autocorrelated to a larger and larger degree as the lags get smaller i.e. the process variation is characterized better and with lower TSE. If the sampling lag is beyond the range, there is no autocorrelation and the data can be considered as a random process on estimating the mean.
- The Nugget Effect; the nugget effect reflects the variance in a 0-D situation i.e. the minimum possible error, including the analytical error. The nugget effect does not have a physical meaning as it is an extrapolated value to zero lag, representing two distinct samples taken at the same location at the same time.

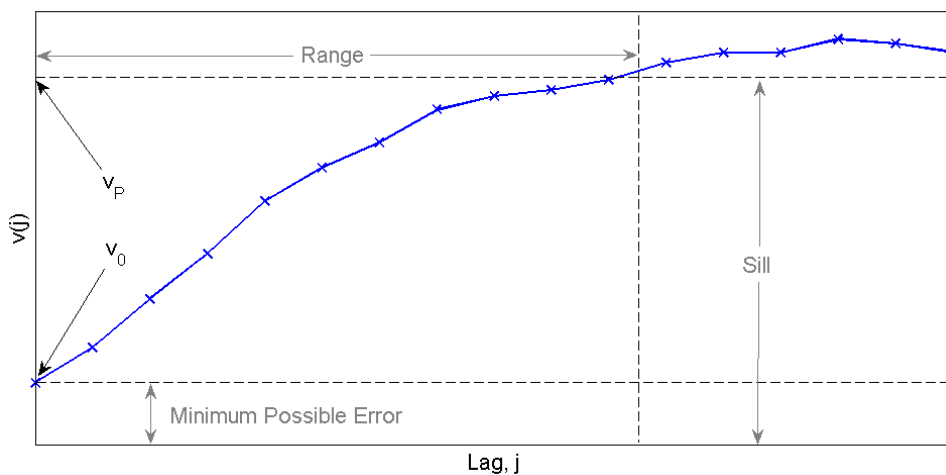


Figure 6. An example of experimental variogram, based on the representation given in (Esbensen, Paoletti & Minkkinen 2012).

It should be noted, that in natural and industrial systems the variogram usually is much more difficult to be interpreted. In such systems, there are always significant scatter when assessing the sill level, the range and the nugget effect in the variogram (Esbensen et al. 2007).

Calculation of the variogram

Variographic analysis requires at least 30 samples taken with a systematic selection mode. The interval should be substantially (at least 1/3) smaller than the assumed adequate routine sampling interval. (Paakkunainen, Reinikainen & Minkkinen 2007) In (Heikka, Minkkinen 1998) it is mentioned that the

variogram is underestimated in the first few multiples of the sampling interval and insignificant for the sampling intervals of five times larger than the experimental one. In summation, a sampling interval less than 1/5 of the expected sampling interval is recommended for the experiments. (Esbensen et al. 2007) mention that the experimental variogram can only be based on one series of maximum of some 40 to 60 samples. In (Minkkinen, Esbensen 2013), it is recommended that number of increments in a variogram should not be less than 60, preferably more than 100. Nevertheless, from the analytical results, the experimental heterogeneity (relative heterogeneity contribution) is first defined as:

$$h_n = \frac{a_n - a_L}{a_L} \cdot \frac{M_n}{\overline{M}_n}$$

Where h_n is the heterogeneity of a sample n , a_n the analytical result and M_n is the size of sample n . \overline{M}_n is the mean sample size and a_L the weighted mean of the lot. If the samples are taken with a cross-stream sample cutter, the sample masses can be weighed. If equal sample volumes are taken, as often in process analysis, or when process analyzers are used, the flow-rate at sampling time should be recorded and used as the sample size. (Paakkunainen, Reinikainen & Minkkinen 2007) The heterogeneity compensates for variation in the fragment masses, which is a major distinction from classical statistics where all units contribute equally (Petersen, Esbensen 2005). The weighted mean of the lot with N samples can be calculated:

$$a_L = \frac{\sum M_n a_n}{\sum M_n} = \frac{1}{N} \sum \frac{M_n}{\overline{M}_n} a_n$$

The experimental variogram, $v(j)$, summarizes the relationship between the mean squared difference between the samples and the lag distance of the corresponding points at which the samples were collected (Esbensen, Paoletti & Minkkinen 2012):

$$v(j) = \frac{1}{2(N-j)} \sum_{n=1}^{N/2} (h_{n+j} - h_j)^2 \approx \frac{1}{2(N-j)a_L^2} \sum_{n=1}^{N/2} (a_{n+j} - a_j)^2$$

The equation shows that the samples can be characterized also by analytical concentrations a_n . In the equations above, the time series $a(t)$ can also be

replaced with a continuous function of space, $a(x)$, as the time and space are conceptually equivalent to their treatment in variographic analysis. (Esbensen, Paoletti & Minkkinen 2012) The lag parameter, j , is described by a division of an inter-sample interval θ and the smallest interval sampled θ_{min} :

$$j = \frac{\theta}{\theta_{min}}$$

Finally, for a complete variogram, also the nugget effect needs to be calculated. In geostatistics, the intercept $\nu(0)$ is usually extrapolated e.g. graphically from the variogram to zero lag. This approach is recommendable especially for periodic processes. Another way defining $\nu(0)$ is to perform a separate experiment with the shortest possible sample interval. The latter method, giving a more reliable estimate, is recommended if short sampling intervals are of interest. If long intervals (e.g. $j > 5$) are to be studied and strong auto-correlation expected, the exact value of $\nu(0)$ is less important. (Heikka, Minkkinen 1997, Paakkunainen, Reinikainen & Minkkinen 2007) The separate experiments should include an adequate amount of samples taken with the highest possible sampling frequency. The estimate for the nugget effect can then be obtained from a (short-range) variogram calculated from this data. (Petersen, Esbensen 2005) Comparison of some methods of estimating the nugget effect value is presented in (Heikka, Minkkinen 1997).

The number of samples in the variographic analysis is restricted, as mentioned above. However, when process measurements are used, the amount of data is huge. The obvious solution is to compress the data by averaging so that the total amount of data points is reduced to an appropriate level. However, a variogram based on a full data set might be useful in ensuring the periodic behavior of the process (Petersen, Esbensen 2005). The periodicity can be confirmed by executing Fourier analysis to the variogram, the heterogeneities and the original data and comparing the resulting periodograms (Paakkunainen, Reinikainen & Minkkinen 2007).

Applying the variographic analysis on estimating the variance estimates for different sampling strategies require some numerical integration. The auxiliary functions related to variographical analysis are presented e.g. in (Gy 2004c, Petersen, Esbensen 2005). The generation of these functions can be based either on point-to-point calculation of the variogram or algebraic modeling.

Variograms in practice

As mentioned before, variographic analysis can be used for determining a suitable sampling interval of systematic sampling. This has been demonstrated e.g. in (Paakkunainen, Reinikainen & Minkkinen 2007, Kohonen, Alatalo & Reinikainen 2012). An example is given in Section 4.1.

It was also mentioned that different sampling strategies (random, stratified, systematic) can be compared using variogram. This has also been done in many studies (Heikka, Minkkinen 1998, Petersen, Esbensen 2005, Paakkunainen, Reinikainen & Minkkinen 2007, Minkkinen, Esbensen 2009, Kohonen, Alatalo & Reinikainen 2012) and one of them is demonstrated in Section 4.1. (Petersen, Esbensen 2005) also demonstrated how many increments should be included in a suitable composite sample in their examples. Composite sampling and variographic analysis has also been combined in (Esbensen et al. 2007, Minkkinen, Esbensen 2009).

(Paakkunainen et al. 2007) applied variograph to a missing sample problem. They studied how reliable the variogram is if the data includes systematic gaps. They also determined how the uncertainty generated by gaps can be estimated when time-averages of auto-correlated time series measurements are reported. Their example came from a simulated data and from combustion processes where the missing gaps of NO_x-data were generated. It was found that random missing gaps in periodic data make the uncertainty estimation very challenging and the most appropriate sampling strategy cannot be chosen based on variogram. If only systematic gaps are present, the problem can be solved by linear interpolation, or more preferably by composite sampling. (Heikka, Minkkinen 1998) suggested that experimental variograms can be calculated also from fewer samples i.e. data collected from normal operation and the variogram can be estimated using cubic smoothing spline function. If this estimated variogram shows some clear changes compared with some earlier, properly executed variographical analysis, it indicates that some process changes have taken place and a full experimental analysis should be conducted.

(Paakkunainen, Reinikainen & Minkkinen 2007) put also an effort to estimate systematic errors with variographic analysis. For this purpose, the variograms of the combustion process measurements and reference measurements were

compared. The behavior of variograms indicated that there was a drift in an analyzer of one of the examples. Both examples also showed a difference between the process measurement and reference measurement, but it cannot be concluded if the difference is due to instrument calibration or sampling system.

A simulation study on the effect of different correct sampling errors is presented in (Minkkinen, Esbensen 2009). The affecting factors simulated were:

- analyte concentration with 5 levels
- heterogeneity type with 7 levels; randomly distributed analyte particles, analyte particles in randomly distributed clusters of 2, 4, 6, 16, and 32 particles, segregated/grading lot
- sample size consisting of different amount of particles with 6 levels
- sampling type (composite sampling, grab sampling)
- sampling strategy (random-, stratified- and systematic sampling)

(Minkkinen, Esbensen 2009) presents the simulator in detail as well as high number of results. The results illustrate the difficulty of comparing and interpreting distributions when the sampling conditions are not fully comparable and specified. They also show the clear advantage of using composite sampling especially when the sample increments are smaller than the cluster size. The simulations also show the effect and relation of GSE and FSE; if an apparent cluster size can be determined, this can be regarded as an effective “particle” and a particle distribution models can be used for estimating FSE. If such determination cannot be made, GSE will be neglected and the models will give underestimated sampling variances.

Variographic analysis can be extended to bilinear projections i.e. PCA, PCR, PLSR scores could be used for variographic characterization instead of individual variables (that might be correlating etc.). (Minkkinen, Esbensen 2013) presents such extension to PCA scores calculated from air quality monitoring data and soil heterogeneity characterization data. They also recommended to plot the time (or spatial) axis simultaneously for the scores and the corresponding variogram for the more comprehensive interpretation of the results. (Kohonen, Alatalo & Reinikainen 2012) have used a multivariate extension of variography to spectral measurements in a semibatch

crystallization process. The variogram of PCA scores showed to be applicable but only for as an indicative tool.

The results of several variographic examples in (Esbensen et al. 2007) demonstrate how composite sampling is more effective on decreasing the total sampling error than optimizing the sampling interval. They studied the effect of outliers in the data and the corresponding variogram. Outliers may influence variograms significantly by masking the true variogram. Therefore it is important to be able to recognize and delete the outliers. (Esbensen et al. 2007) also showed the importance of pre-treating the original data series by de-trending or time-segmentation of the data in order to reveal the hidden data structures through variograms. This kind of pre-treatment corresponds with the geostatistical prerequisite of stationarity. They also suggested that both the original data and the pre-treated data should be analyzed with variography simultaneously.

The variogram was efficiently used as a process analytical tool (PAT) in (Esbensen et al. 2007). An example given from a biogas process monitoring revealed a trend of seven days and 21 days. The former one was related to weekends, when a constant pre-mix of the feedstock cannot be guaranteed. The latter one provided new process insight as it was connected with the raw material logistics from certain suppliers. Variography was used for a corporate quality control (QC) in an example given in (Esbensen, Mortensen 2009). Six manufacturing plants of the same corporate aimed to produce the same product with similar properties. Variographical analysis revealed distinctions between the processes which were assumed to operate with equal reliably and efficiency. Three plants with high sill levels were assigned to take a complete TOS analysis. One plant also had to take actions against the periodic variability observed in the analysis.

Heikka and Minkkinen have aimed at estimating the limiting value of the variogram (the nugget effect) more precisely with spline functions (Heikka, Minkkinen 1997, Heikka, Minkkinen 1998). They also found in their analysis of pulp bleaching plant data that the limiting values of the variograms were high due to periodicity and random noise. Additionally, they concluded that if the variogram increases slowly, more precise laboratory measurements are beneficial on the process monitoring and controlling point of view. If the

variogram increases quickly, the control should rely only on the quick process analyzers.

There exists another theory related to variographic analysis. (Heikka 1996) compares the Gy's sampling theory and the one introduced by Kateman and Müskens. The latter theory is based on the autocorrelation analysis and equations derived from properties of stationary, stochastic processes with normally distributed properties, hence corresponding to the non-periodic continuous component of Gy's integration error (Heikka 1996).

4. Applications

4.1. TOS and variographic analysis

TOS and variographic analysis was first developed into soil science and mining operations. The theories belong to a branch of statistics focusing on spatial and spatiotemporal datasets called geostatistics. Nowadays, geostatistics have been applied in a large number of studies in various fields. In this literature review, the focus is on variographic analysis in process- and environmental engineering related applications and process measurements. The references cited in this review involved the following applications (some of those also outside of process engineering applications):

- Chlorine dioxide and chlorite residuals in pulp bleaching and variographic analysis (Heikka, Minkkinen 1997, Heikka, Minkkinen 1998)
- Sulfur discharge in wastewater and variographic analysis (Paakkunainen, Reinikainen & Minkkinen 2007)
- NO_x emissions from a combustion process and variographic analysis (Paakkunainen et al. 2007), (Paakkunainen, Reinikainen & Minkkinen 2007)
- CO measurements from a combustion process and variographic analysis (Paakkunainen et al. 2007)
- Variance estimation of the sulfur emission sources in a pulp mill and variographic analysis (Minkkinen 2004)
- Yield (CH₄) and pollutant (H₂S) level monitoring of a biogas plant and variographic analysis (Esbensen et al. 2007)
- Trade prices of zink and spot prices for crude oil and variographical analysis (Esbensen et al. 2007)
- Flotation plant feed and variographic analysis (Esbensen et al. 2007)
- Online pressure, temperature and oxygen concentration of the feed water in a power plant and variographic analysis (Petersen, Esbensen 2005)
- Spectral instruments in semibatch chrystallization process and variographic analysis (Kohonen, Alatalo & Reinikainen 2012)
- Air quality monitoring and variographic analysis (Minkkinen, Esbensen 2013)
- Soil heterogeneity characterization and variographic analysis (Minkkinen, Esbensen 2013)

- Uranium content of the feed in a uranium ore processing plant and variographic analysis (Gy 2004c)
- Feed to a cement kiln and variographic analysis (Gy 2004c)
- Chromite concentration estimation from SEM images and TOS (Korpelainen et al. 2002)
- Sampling procedures for CO₂ emissions estimation in a coal power plant and TOS (Wagner, Esbensen 2011)

Two examples are represented here in more detail. The first one illustrates how variographic analysis can be used for determining a sampling plan. The second one shows how the periodicity of the data affects the selection of sampling strategy.

(Paakkunainen, Reinikainen & Minkkinen 2007) uses variographic analysis to improve the estimation of the confidence interval of sulfur discharge in wastewater over a one-year period. For the analysis, they collected one water sample daily for a one-month period. The heterogeneity data from (Paakkunainen, Reinikainen & Minkkinen 2007) is reinterpreted in Figure 7 and the corresponding variogram, as well as the relative standard deviations for two different sampling strategies are calculated. The standard deviation of the heterogeneities i.e. the relative standard deviation of the process is 28.2%. The variogram shows no periodicity and therefore the interpretation of it is straightforward. If one sample in a week is collected (i.e. 7 day sampling interval, 52 samples in a year), the relative standard deviation with systematic sampling is 7.8%. This results as a relative standard deviation of the annual sulfur discharge of $0.078/\sqrt{52}=1.1\%$. If the sampling interval is halved (3.5 day sampling interval, 104 samples in a year), the relative standard deviation with systematic sampling is 5.8% and the corresponding uncertainty in the annual estimate would be $0.058/\sqrt{104}=0.57\%$. However, if one assumes that the results follow a normal distribution, random sampling is used and an acceptable relative standard deviation is 1.1%, the required number of samples would be $0.282^2/0.011^2=657$.

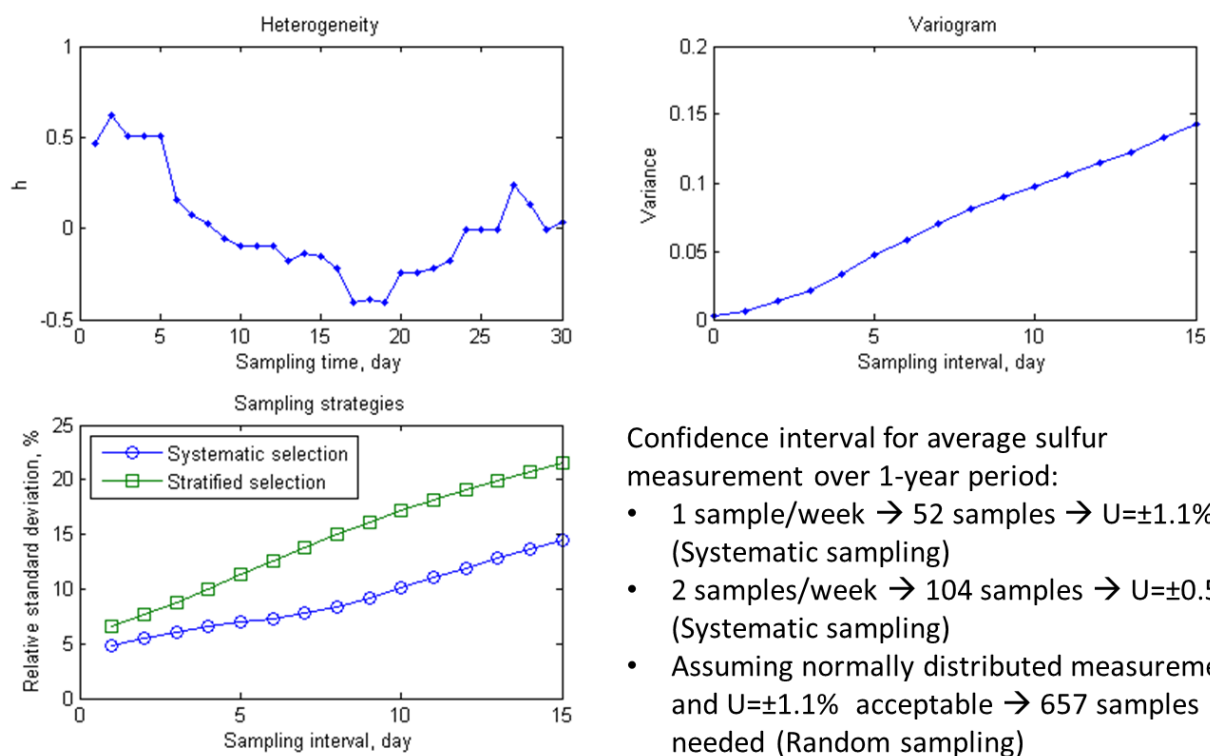


Figure 7. Variographical analysis of sulfur discharge. Heterogeneity data interpreted from (Paakkunainen, Reinikainen & Minkkinen 2007).

The second example is a simulated example originally presented in (Petersen, Esbensen 2005). In this variographic experiment, samples are taken every minute for a one hour period and the concentrations of one component are analyzed. The variogram in Figure 8 indicates a cyclic behavior with a period around 13 minutes, a behavior that cannot be seen in the analytical results. Sampling with such a period should be avoided. The auxiliary functions are utilized to provide the variances and standard deviations of the sampling with different sampling strategies. The x-axis is reversed in this presentation in order to show the number of increments in a composite sample and the corresponding variance (or standard deviation). The analysis shows that instead randomly sampling the process once an hour, using a composite sample of 10 increments decreases the absolute standard deviation from a value higher than 3 to a value near 0.5. This relates to sampling the process every 6 minutes. Due to the cyclic behavior of the process, the stratified random sampling is preferable.

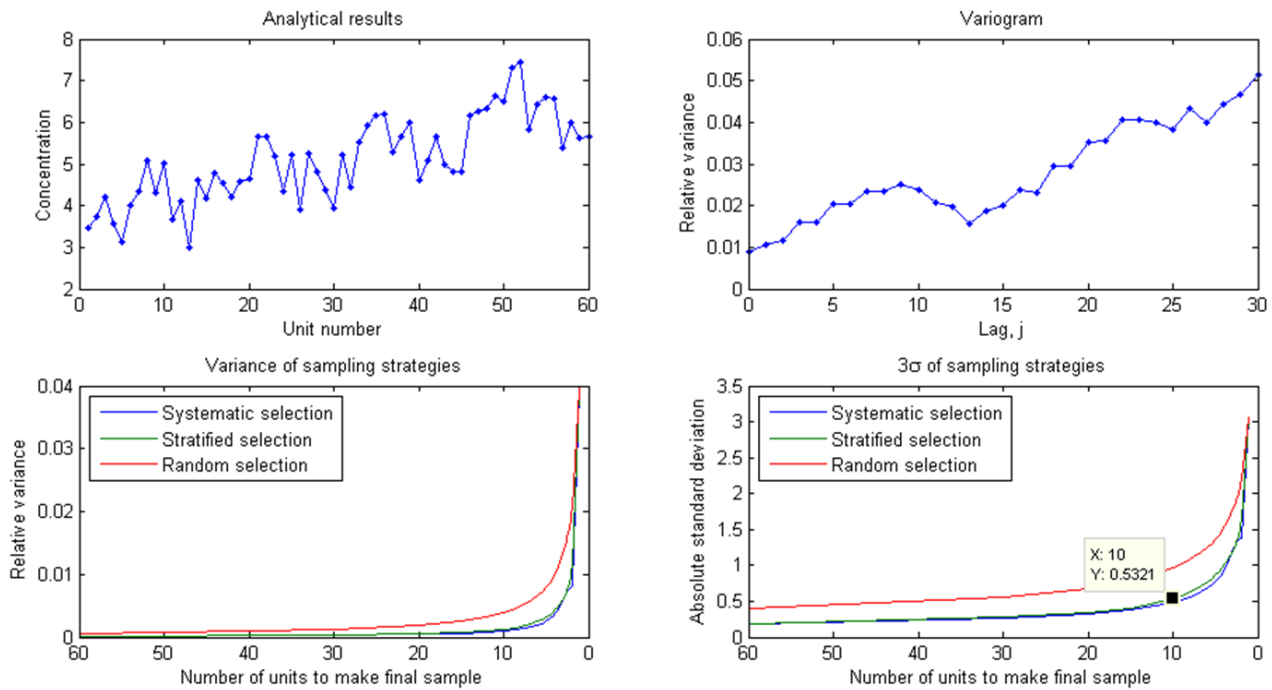


Figure 8. Variographical analysis of simulated data. Analytical results interpreted from (Petersen, Esbensen 2005).

4.2. Uncertainty estimation for virtual instruments

Extending the uncertainty estimation into automation systems requires studying e.g. the uncertainty components in an automation system and in the data transfer network. A practical example has been given in (Koivisto 2003). More generally, the virtual measurement can include a large amount of uncertainty sources. A comprehensive metrological characterization of the virtual measurement instrument is given in (Caldara, Nuccio & Spataro 2000) and (Nuccio, Spataro 2002). The problem may be decomposed in a following way:

- Transducers and signal conditioning accessories; their error are predominant in comparison with other sources, but also case-sensitive i.e. no general treatment can be made, but separate analysis of particulate situations is required (Nuccio, Spataro 2002).
- Acquisition board (sampler, analog-to-digital conversion, clock generator); the systematic and random uncertainty sources are presented in Figure 9.
- General-purpose computer
- Software (data acquisition board control, digital signal processing, user interface); errors arise from bias of processing algorithms, rounding phenomena (finite wordlength)

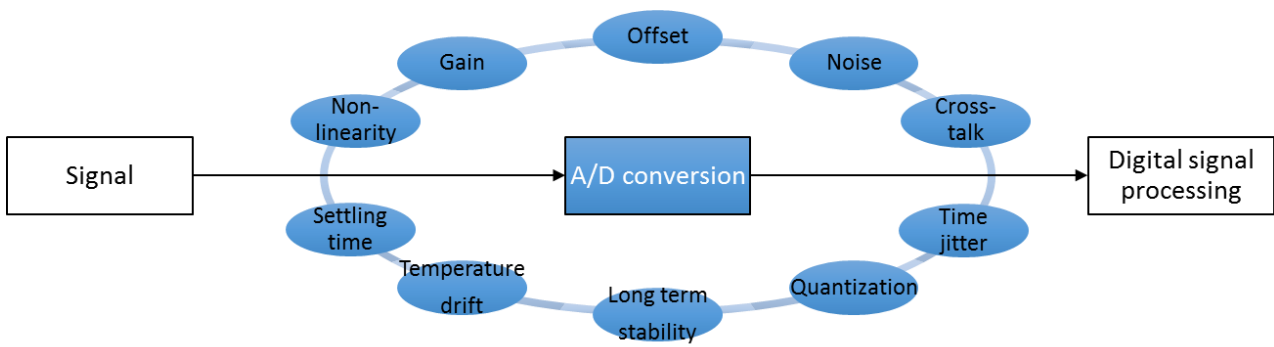


Figure 9. Uncertainty sources of the A/D conversion process, redrawn from (Nuccio, Spataro 2002).

The A/D-conversion comprises many uncertainty sources, as presented in Figure 9. The temperature drift in that Figure refers to the temperature drift of the onboard calibration reference. Pre-gain offset, gain and post-gain offset are all affected by their temperature drifts. Some of the uncertainties are usually provided as a Type B uncertainty by the manufacturers, but defining the others may require using Monte Carlo Simulation. (Caldara, Nuccio & Spataro 2000) gives a simulated example of uncertainty estimation of a virtual instrument. (Nuccio, Spataro 2002) compares the uncertainty estimates gained from their theoretical approach, from a numerical method and from an experimental method. The A/D-conversion and its uncertainty have also been discussed in (Clemens 2000), where the typical random errors and A/D errors were combined non-additively.

The uncertainties arising from the software, i.e. from digital signal elaboration algorithms, are treated in (Betta, Liguori & Pietrosanto 1999). They proposed a white box-approach to the uncertainty estimation of the output signal in time-domain. In (Betta, Liguori & Pietrosanto 2000), the authors extended the uncertainty estimation to DFT-based (discrete Fourier transform) instruments. (De Santo et al. 2004) have made uncertainty estimation to digital images using the abovementioned white-box approach. In their study, (De Santo et al. 2004) decomposed the image-based measurement problem into image acquisition, image processing and measurement extraction.

(Locci, Muscas & Ghiani 2002) suggest an auto-evaluation of bias and uncertainty of a virtual instrument. They consider the uncertainties arising from the data acquisition and A/D-conversion system including quantization, noise,

offset and gain error. They also make a numerical comparison of GUM, MCM and mathematical evaluation of the central moments of the random variables.

Evaluation of the uncertainty and fit-for-purpose (FFP) of the on-site measurements is presented in [Boon & Ramsey 2012]. On-site measurements have similarities to intelligent measurements as they take an advantage of more cost-efficient sampling by increasing the number of samples and also the disadvantages are similar, e.g. less-sensitive detection limits. FFP analysis, presenting also one solution to the compositing of samples, is presented in the next Section.

5. Fit-for-purpose

A sampling plan is not only guided by the uncertainty of the measurement, but also by the costs of the measurement. Naturally, minimizing the costs is an intriguing option for economic point-of-view. Therefore, if a certain uncertainty limit exists, or is set by legislation, the sampling plan can be optimized by minimizing the costs with the boundary condition to fulfill the limits. However, in decision making, the optimization of the sampling plan is more difficult. Some level of precision is of course required, but it is more important to be able to have a measurement result (estimate + uncertainty) that can be used as a basis of a reliable decision making. Naturally, time is valuable in decision making, so sometimes it is beneficial to have more samples in less time, although the uncertainty is higher. Therefore, the loss function should take into account the costs of the investigation and the possible losses due to misclassification as a function of the measurement uncertainty (Ramsey, Boon 2012). Higher uncertainty could, for example, lead to a wider threshold for action levels.

This kind of evaluation of the measurement on its ultimate purpose is one of the three approaches listed for judging fitness-for-purpose (FFP) of measurements using uncertainty (Ramsey, Ellison 2007). The two other are: 1) Setting a maximum value of uncertainty considered acceptable and 2) Comparing the measurement variance to the variance of the measurements between the different sampling targets; e.g. locating targets with significantly higher concentrations, indicated by the low impact of measurement variance on the total variance. However, considering the ultimate purpose is the most generally applicable. This is the FFP considered in this review. It should be noted that in a wider sense, the FFP of analytical methods may consist of characteristics such as selectivity, repeatability, linearity, sensitivity, limit of detections, robustness, trueness.

(Ramsey, Boon 2012) compared the FFP of in-situ and ex-situ measurements in geochemical measurements, where the ex-situ laboratory measurements may take even days with several sample handling and mass reduction steps. They listed some of the pros and cons of in-situ and ex-situ measurements, which are presented in Table 2. The random component was estimated using the duplicate method, the systematic component was estimated using certified reference materials and the bias between the measurements was modeled using maximum likelihood estimation. The FFP evaluation was based on the

OCLI method (optimized contaminated land investigation), that uses a cost function based upon not just the cost of taking the measurements (sampling and chemical analysis), but also on the costs arising from the consequences of misclassification of two types. The first one, the false positive, represents a case where the land is erroneously classified as contaminated when it is not. The second one, the false negative, classifies the land erroneously as uncontaminated. Both misclassifications may lead to extra costs as land might be remediated needlessly, contamination left in place may be detected later on lead to delays in site development or contamination has human health effects, or subsequent litigation.

In (Boon, Ramsey 2012), FFP of in-situ and ex-situ geochemical measurements was determined and the measurement method was then adjusted to really achieve FFP. The methods were similar to those in (Ramsey, Boon 2012). The adjusting was made by increasing the mass i.e. the number of increments in the composite samples. Assuming a representative sampling, the dependence on the sampling variance and number of increments is:

$$\frac{n_2}{n_1} = \left(\frac{s_1}{s_2}\right)^2 \Leftrightarrow n_2 = \left(\frac{s_1}{s_2}\right)^2 n_1$$

Where n_1 and n_2 are the number of increments in the composite sample and s_1 and s_2 are the standard deviations of those composite samples. The optimal uncertainty for FFP measurement, s_2 , is attained from the RANOVA analysis. It should be noted, that if both the sampling uncertainty and analytical uncertainty are important (i.e. at comparable level), adjusting the sampling plan has a smaller effect on the total uncertainty and s_2 used in equation should be determined as $s_{2,new} = \sqrt{s_{2,old}^2 - s_{analysis}^2}$.

In their first case study, (Boon, Ramsey 2012) found that the on-site measurements comprising of three increments would be FFP for avoiding false-positive classification. However, avoiding false-negative classification in an optimal manner (FFP) was found to be impractical, as the composite sample should include 147 increments. In the second case study, they found that 14 increments should be taken instead of 5 increments in the composite sample in order to achieve FFP measurement for avoiding false positives. The original sampling plan was already more than FFP for avoiding false negatives, so the

increased amount of increments in the composite sample would also be more than FFP.

Table 2. Comparison of in-situ and ex-situ measurements (Ramsey, Boon 2012).

	Pros	Cons
In-situ	<ul style="list-style-type: none"> + Lower cost + Accuracy + Immediate estimates + Higher sampling densities + Less sample preparation + Less storage and waste + Spatial heterogeneity preserved 	<ul style="list-style-type: none"> - Less reliable - Resolution of detection limits - Training
Ex-situ	<ul style="list-style-type: none"> + Reproducibility + Traceability + Quality control 	<ul style="list-style-type: none"> - Losses of analyte in mass reduction and processing

Instead of using time-series analysis for the variance estimation, the fit-for-purpose evaluation can also be based on the variance estimates from variographic analysis. In (Minkinen 2004), examples of optimizing a nested sampling plan are given. The optimization can be made with respect to the costs of the investigation for preset variance, or with respect to the minimum variance for preset costs. A numerical example of the former one concerns with the determination of cobalt in nickel cathodes (equal strata). A numerical example of the latter one optimizes the allocations of emission control measurements to estimate the sulfur balance of a pulp mill (unequal strata). The results then recommend the optimized sampling protocols (sampling frequency and allocation, sample masses).

6. References

1. Betta, G., Liguori, C. & Pietrosanto, A. 1999, "Structured approach to estimate the measurement uncertainty in digital signal elaboration algorithms", *Science, Measurement and Technology, IEE Proceedings -*, vol. 146, no. 1, pp. 21-26.
2. Betta, G., Liguori, C. & Pietrosanto, A. 2000, "Propagation of uncertainty in a discrete Fourier transform algorithm", *Measurement*, vol. 27, no. 4, pp. 231-239.
3. Boon, K.A. & Ramsey, M.H. 2012, "Judging the fitness of on-site measurements by their uncertainty, including the contribution from sampling", *Science of The Total Environment*, vol. 419, no. 0, pp. 196-207.
4. Caldara, S., Nuccio, S. & Spataro, C. 2000, "Measurement uncertainty estimation of a virtual instrument", *Instrumentation and Measurement Technology Conference, 2000. IMTC 2000. Proceedings of the 17th IEEE*, pp. 1506.
5. Clemens, E. 2000, "Evaluation of measurement uncertainty in the presence of combined random and analogue-to-digital conversion error", *Meas. Sci. Technol.*, vol. 11, pp. 1359-1363.
6. Cox, M., Harris, P. & Siebert, B.R.-. 2003, "Evaluation of Measurement Uncertainty Based on the Propagation of Distributions Using Monte Carlo Simulation", *Measurement Techniques*, vol. 46, no. 9, pp. 824-833.
7. da Silva Hack, P. & ten Caten, C.S. 2012, "Measurement Uncertainty: Literature Review and Research Trends", *Instrumentation and Measurement, IEEE Transactions on*, vol. 61, no. 8, pp. 2116-2124.
8. De Santo, M., Liguori, C., Paolillo, A. & Pietrosanto, A. 2004, "Standard uncertainty evaluation in image-based measurements", *Measurement*, vol. 36, no. 3-4, pp. 347-358.
9. Ellison, S.L.R. & Williams, A. (eds) 2012, *Eurachem/CITAC Guide: Quantifying Uncertainty in Analytical Measurement*, Third edition edn, Eurachem.
10. Esbensen, K.H., Friis-Petersen, H.H., Petersen, L., Holm-Nielsen, J.B. & Mortensen, P.P. 2007, "Representative process sampling — in practice:

Variographic analysis and estimation of total sampling errors (TSE)", *Chemometrics and Intelligent Laboratory Systems*, vol. 88, no. 1, pp. 41-59.

11. Esbensen, K.H. & Mortensen, P.P. 2009, "Chapter 3 Process Sampling: Theory of Sampling - the Missing Link in Process Analytical Technologies (PAT)" in *Process Analytical Technology*, ed. K.A. Bakeev, 2nd edn, Wiley, , pp. 37-80.
12. Esbensen, K.H., Paoletti, C. & Minkkinen, P. 2012, "Representative sampling of large kernel lots I. Theory of Sampling and variographic analysis", *TrAC Trends in Analytical Chemistry*, vol. 32, no. 0, pp. 154-164.
13. Ferrero, A., Gamba, R. & Salicone, S. 2004, "A method based on random-fuzzy variables for online estimation of the measurement uncertainty of DSP-based instruments", *Instrumentation and Measurement, IEEE Transactions on*, vol. 53, no. 5, pp. 1362-1369.
14. Ferrero, A. & Salicone, S. 2004, "The random-fuzzy variables: a new approach to the expression of uncertainty in measurement", *Instrumentation and Measurement, IEEE Transactions on*, vol. 53, no. 5, pp. 1370-1377.
15. Ferrero, A. & Salicone, S. 2003, "An innovative approach to the determination of uncertainty in measurements based on fuzzy variables", *Instrumentation and Measurement, IEEE Transactions on*, vol. 52, no. 4, pp. 1174-1181.
16. Gy, P. 2004a, "Part I: Sampling of discrete materials—a new introduction to the theory of sampling: I. Qualitative approach", *Chemometrics and Intelligent Laboratory Systems*, vol. 74, no. 1, pp. 7-24.
17. Gy, P. 2004b, "Part II: Sampling of discrete materials: II. Quantitative approach—sampling of zero-dimensional objects", *Chemometrics and Intelligent Laboratory Systems*, vol. 74, no. 1, pp. 25-38.
18. Gy, P. 2004c, "Part III: Sampling of discrete materials: III. Quantitative approach—sampling of one-dimensional objects", *Chemometrics and Intelligent Laboratory Systems*, vol. 74, no. 1, pp. 39-47.

19. Gy, P. 2004d, "Part IV: 50 years of sampling theory—a personal history", *Chemometrics and Intelligent Laboratory Systems*, vol. 74, no. 1, pp. 49-60.
20. Gy, P. 2004e, "Part V: Annotated literature compilation of Pierre Gy", *Chemometrics and Intelligent Laboratory Systems*, vol. 74, no. 1, pp. 61-70.
21. Heikka, R. 1996, "Comparison of two sampling theories to calculate the integration error for one-dimensional processes", *Chemometrics and Intelligent Laboratory Systems*, vol. 33, no. 2, pp. 147-157.
22. Heikka, R. & Minkkinen, P. 1998, "Integration of the variogram using spline functions for sampling error estimation", *Chemometrics and Intelligent Laboratory Systems*, vol. 44, no. 1–2, pp. 205-211.
23. Heikka, R. & Minkkinen, P. 1997, "Comparison of some methods to estimate the limiting value of the variogram, $\gamma(h)$, for the sampling interval $h=0$ in sampling error estimation", *Analytica Chimica Acta*, vol. 346, no. 3, pp. 277-283.
24. Heinonen, M. 2010, *Lecture material: Introduction to Metrology*.
25. Herrador, M.Á, Asuero, A.G. & González, A.G. 2005, "Estimation of the uncertainty of indirect measurements from the propagation of distributions by using the Monte-Carlo method: An overview", *Chemometrics and Intelligent Laboratory Systems*, vol. 79, no. 1–2, pp. 115-122.
26. Hessling, J.P. 2009, "A novel method of evaluating dynamic measurement uncertainty utilizing digital filters", *Measurement Science and Technology*, vol. 20, no. 5, pp. 055106.
27. JCGM100 2008, *Evaluation of measurement data — Guide to the expression of uncertainty in measurement*, Joint Committee for Guides in Metrology.
28. JCGM101 2008, *Evaluation of measurement data – Supplement 1 to the "Guide to the expression of uncertainty in measurement" – Propagation of distributions using a Monte Carlo method*, Joint Committee for Guides in Metrology.

29. Kacker, R. & Jones, A. 2003, "On use of Bayesian statistics to make the Guide to the Expression of Uncertainty in Measurement consistent", *Metrologia*, vol. 40, pp. 235-248.
30. Kohonen, J., Alatalo, H. & Reinikainen, S. 2012, "Evaluation of variation in dynamic processes via online spectrometers", *Journal of Chemometrics*, vol. 26, no. 6, pp. 333-339.
31. Koivisto, A. 2003, *Reliability of pulp yield prediction estimate for TMP-line (in Finnish)*, University of Oulu.
32. Korpelainen, M., Reinikainen, S.-., Laukkanen, J. & Minkkinen, P. 2002, "Estimation of uncertainty of concentration estimates obtained by image analysis", *Journal of Chemometrics*, vol. 16, no. 8-10, pp. 548-554.
33. Kurfürst, U., Desaulles, A., Rehnert, A. & Muntau, H. 2004, "Estimation of measurement uncertainty by the budget approach for heavy metal content in soils under different land use", *Accreditation and Quality Assurance*, vol. 9, no. 1-2, pp. 64-75.
34. Lambkin, D., Nortcliff, S. & White, T. 2004, "The importance of precision in sampling sludges, biowastes and treated soils in a regulatory framework", *TrAC Trends in Analytical Chemistry*, vol. 23, no. 10–11, pp. 704-715.
35. Locci, N., Muscas, C. & Ghiani, E. 2002, "Evaluation of uncertainty in measurements based on digitized data", *Measurement*, vol. 32, no. 4, pp. 265-272.
36. Lovett, T.E., Ponci, F. & Monti, A. 2006, "A polynomial chaos approach to measurement uncertainty", *Instrumentation and Measurement, IEEE Transactions on*, vol. 55, no. 3, pp. 729-736.
37. Lyn, J.A., Ramsey, M.H., Fussell, R.J. & Wood, R. 2003, "Measurement uncertainty from physical sample preparation: Estimation including systematic error", *Analyst*, vol. 128, no. 11, pp. 1391-1398.
38. Lyn, J.A., Ramsey, M.H., Damant, A.P. & Wood, R. 2007, "Empirical versus modelling approaches to the estimation of measurement uncertainty caused by primary sampling", *Analyst*, vol. 132, no. 12, pp. 1231-1237.

39. Mauris, G., Lasserre, V. & Foulloy, L. 2001, "A fuzzy approach for the expression of uncertainty in measurement", *Measurement*, vol. 29, no. 3, pp. 165-177.
40. Minkkinen, P. 2004, "Practical applications of sampling theory", *Chemometrics and Intelligent Laboratory Systems*, vol. 74, no. 1, pp. 85-94.
41. Minkkinen, P.O. & Esbensen, K.H. 2009, "Grab vs. composite sampling of particulate materials with significant spatial heterogeneity—A simulation study of “correct sampling errors”", *Analytica Chimica Acta*, vol. 653, no. 1, pp. 59-70.
42. Minkkinen, P. & Esbensen, K.H. 2013, "Multivariate variographic versus bilinear data modeling", *Journal of Chemometrics*, , pp. n/a-n/a.
43. Nuccio, S. & Spataro, C. 2002, "Approaches to evaluate the virtual instrumentation measurement uncertainties", *Instrumentation and Measurement, IEEE Transactions on*, vol. 51, no. 6, pp. 1347-1352.
44. O'Donnell, G.E. & Hibbert, D.B. 2005, "Treatment of bias in estimating measurement uncertainty", *Analyst*, vol. 130, no. 5, pp. 721-729.
45. Paakkunainen, M., Reinikainen, S.-. & Minkkinen, P. 2007, "Estimation of the variance of sampling of process analytical and environmental emissions measurements", *Chemometrics and Intelligent Laboratory Systems*, vol. 88, no. 1, pp. 26-34.
46. Paakkunainen, M., Kilpeläinen, J., Reinikainen, S. & Minkkinen, P. 2007, "Effect of missing values in estimation of mean of auto-correlated measurement series", *Analytica Chimica Acta*, vol. 595, no. 1–2, pp. 209-215.
47. Petersen, L. & Esbensen, K.H. 2005, "Representative process sampling for reliable data analysis? a tutorial", *Journal of Chemometrics*, vol. 19, no. 11-12, pp. 625-647.
48. Petersen, L., Minkkinen, P. & Esbensen, K.H. 2005, "Representative sampling for reliable data analysis: Theory of Sampling", *Chemometrics and Intelligent Laboratory Systems*, vol. 77, no. 1–2, pp. 261-277.
49. Ramsey, M.H. & Ellison, S.L.R. (eds) 2007, *Eurachem/EUROLAB/CITAC/Nordtest/AMC Guide: Measurement*

uncertainty arising from sampling: a guide to methods and approaches, Eurachem.

50. Ramsey, M.H. & Boon, K.A. 2012, "Can in situ geochemical measurements be more fit-for-purpose than those made ex situ?", *Applied Geochemistry*, vol. 27, no. 5, pp. 969-976.
51. Reznik, L. & Dabke, K.P. 2004, "Measurement models: application of intelligent methods", *Measurement*, vol. 35, no. 1, pp. 47-58.
52. Salicone, S. 2013, "The theory of evidence: a new promising approach to the evaluation and expression of measurement uncertainty", *Instrumentation & Measurement Magazine, IEEE*, vol. 16, no. 1, pp. 18-23.
53. Urbanski, M.K. & Wąsowski, J. 2003, "Fuzzy approach to the theory of measurement inexactness", *Measurement*, vol. 34, no. 1, pp. 67-74.
54. Wagner, C. & Esbensen, K.H. 2011, "A systematic approach to assessing measurement uncertainty for CO₂ emissions from coal-fired power plants – Missing contributions from the Theory of Sampling (TOS)", *Chemical Engineering Research and Design*, vol. 89, no. 9, pp. 1572-1586.