



Review

Theory of sampling (TOS) *versus* measurement uncertainty (MU) – A call for integrationKim H. Esbensen^{a,b}, Claas Wagner^{b,*}^a Geological Survey of Denmark and Greenland, Copenhagen, Denmark^b ACABS Research Group, Aalborg University, Campus Esbjerg (AAUE), Denmark

ARTICLE INFO

Keywords:

Correct sampling error
EURACHEM/CITAC
GUM
Inconstant sampling bias
Incorrect sampling error
Measurement uncertainty
Sampling bias
Sampling error
Theory of sampling
Total sampling error

ABSTRACT

We assess current approaches to measurement uncertainty (MU) with respect to the complete ensemble of sources affecting the measurement process, in particular the extent to which sampling errors as set out in the Theory of Sampling (TOS) are appropriately considered in the GUM and EURACHEM/CITAC guides. All pre-analysis sampling steps play an important, often dominant role in the total uncertainty budget, thereby critically affecting the validity of MU estimates, but most of these contributions are not included in the current MU framework. The TOS constitutes the only complete theoretical platform for dealing appropriately with the entire pathway from field sample to test portion. We here propose a way to reconcile the often strongly felt differences between MU and TOS. There is no need to debate terminology, as both TOS and MU can be left with their current usages.

© 2014 Elsevier Ltd. All rights reserved.

Contents

1. Introduction	93
2. TOS pathway	94
2.1. Lot dimensionality	94
2.2. Heterogeneity	95
2.3. The sampling process	97
2.4. Sampling errors – and their effects	97
2.5. Sampling in practice	99
2.6. TOS summary	99
3. MU pathway	99
4. TOS – the missing link: a call for integration	100
5. Conclusion	103
6. Postscript: terminology issues	103
7. Explanation 1. Heterogeneity	103
8. Explanation 2. Intrinsic uncertainty	104
9. Explanation 3. Gross errors <i>versus</i> incorrect sampling errors	104
10. Explanation 4. Systematic measurement error <i>versus</i> sampling bias	104
11. Explanation 5. Certified reference materials (CRMs)	105
12. Explanation 6. RST (reference sampling target), SPT (sampling proficiency test), and CTS (collaborative trial in sampling)	105
13. Explanation 7. Variographics (TOS)	105
References	106

1. Introduction

The purpose of sampling is to extract an amount of material from a 'lot' (also termed the 'sampling target'), which can be

* Corresponding author. Tel.: +49 4131 36920.

E-mail address: cw@bio.aau.dk (C. Wagner).

documented to be representative of the lot. It is evident that sampling should be optimized before analysis, as it is always preceding, and no longer has any interaction with the test portion when this is analyzed. However, a non-representative sampling process will always deliver an invalid aliquot for MU characterization. A specific sampling process can either be representative (full definition below), or not; only the first results in representative samples, the latter results in mass-reduced undefined lumps of material without provenance ('specimens' in Theory of Sampling (TOS) parlance). Only analytical results pertaining to representative aliquots reduce the measurement uncertainty of the full sampling-and-analysis process to its minimum (based on an analytical process in full control). Sampling process correctness (full definition below) and representativity are therefore core elements of concern to both the sampling process and for minimum measurement uncertainty objectives in analysis.

The TOS has been established in the past 60 years as the only theoretical framework that deals in full with sampling, the representativity concept and all practical aspects involved in achieving the required representative test portion. The full pathway 'from-lot-to-analytical-aliquot' is complex, and, in some aspects, counter-intuitive due to the phenomenon of heterogeneity, and it is subject to many types of uncertainty contributions not only concerning analysis. GUM [1] and the EURACHEM/CITAC [2] guide focus on estimating the total measurement uncertainty (GUM and the EURACHEM/CITAC guide are termed 'MU' in the following). There is a subtle, but far from trivial, distinction in perspectives: the TOS focuses on the conceptual and practical active steps needed for minimizing all sampling contributions to MU, while MU focuses on passive estimation of the total MU of the sampling-plus-analysis process irrespective of its magnitude, as based on the test portion. However, if the test portion is not representative [i.e. if all sampling error effects have not been reduced as appropriate (full description below)], all MU estimates are compromised in that they will always be structurally too low (and always to an unknown degree). Both frameworks are in their nature significantly complex but do not overlap if and when treated with the conceptual clarity illustrated in Fig. 1.

This study points out the main discrepancies between TOS and MU and presents reasons why there is a strong need for reconciliation and how this can easily be achieved.

For readers not well versed in the TOS, a comprehensive introduction can be found [3–11]. For further insight into MU, regarding aspects not treated in this article, the reader is referred to GUM [1] and the EURACHEM/CITAC guide [2].

Fig. 2 gives a structural overview of the line of argumentation elaborated in the sections below. The starting point of every measurement process is the primary lot, or the sampling target, which is defined in the appropriate sections below. All lots are characterized by significant material heterogeneity, a concept only fully defined by the TOS, and here crucially sub-divided into constitutional heterogeneity (CH) and distributional heterogeneity (DH).

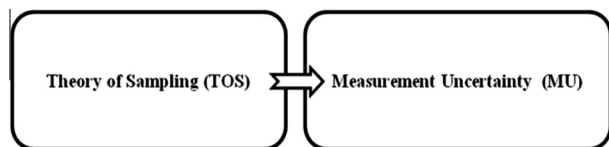


Fig. 1. Non-overlapping, interconnected disciplines: Theory of Sampling (TOS) versus Measurement Uncertainty (MU). The responsibility of the TOS is to deliver a representative analytical aliquot (arrow) for analysis with documentable minimum total sampling errors (TSEs) because of competent command of the entire lot-to-aliquot sampling process, while all errors characterizing the analytical processes (TAE) are validated by a comprehensive MU_{analysis} estimation. Both disciplines are needed; indeed, they complement one another completely.

The heterogeneity concept and its many manifestations are further introduced in detail below.

The pathway, from sampling target to MU, and its implicit estimate of the sampling-process-error effects (MU_{sampling}) is presented in the upper part of Fig. 2, and explained and argued in Section 3. The TOS pathway is depicted in the lower part of Fig. 2 and described in the following section. The concept of MU related to analytical measurement (MU_{analysis}) is referred to in both MU and TOS pathways, although termed 'Total Analytical Error' (TAE) in the TOS.

There is no need for worry about possible confusion stemming from the different terminologies in these two approaches; this state of affairs is unavoidable, since it evolved in two distinct scientific communities with very little interaction (so far). By analyzing the existing, crucial differences, we reach the conclusion that a call for structured reconciliation is timely and mutually beneficial, and that there is not much danger of a terminology debacle, as both TOS and MU can be left with their current usages.

2. TOS pathway

2.1. Lot dimensionality

Following the pathway in Fig. 2 (from left to right) both terminologies 'lot' (TOS) and 'sampling target' (MU) recognize that the extracted material portion (the 'primary sample'), which will eventually be mass reduced to the analytical aliquot, must be representative of the lot. The lot refers to the physical and geometrical aspect of the sampling target (e.g., material on the conveyer belt, or in stockpiles, shiploads, or natural systems).

In the TOS, lot dimensionality is characterized by specifying the operative number of dimensions to be 'covered' during the sampling process, defining one-, two- and three-dimensional (1-D, 2-D and 3-D) lots and the special case of a zero-dimensional (0-D) lot, reflecting the effective number of dimensions involved in the sampling process. (A 0-D lot refers to a lot that can be effectively, mixed, moved and sampled throughout with complete correctness (see below). Usually, these are comparatively small lots, which can easily be manipulated). The concept of lot dimensionality becomes clear, e.g., when considering an elongated material stream, as is the case for material on conveyer belts.

According to the MU definition, this sampling target should be termed 1-D, since one dimension of the physical geometrical aspect dominates, while, according to the TOS, it is essential to consider the applied sampling method as interacting with an effective number of dimensions during the sampling process. Employing grab sampling (full definition below) on such an elongated material stream, a widely-applied, but fundamentally-flawed, extraction method, would make this lot effectively 3-D (and not 1-D), since singular grab samples are most likely only taken from the top surface part of the moving material flux, so far from covering both the transverse lot dimensions fully (i.e. width and thickness). By contrast, a cross-stream cutter (a sampling device especially designed for elongated material fluxes) would cover the entire depth and width of the stream, thereby fully reducing the sampling lot to one dimension (i.e. the longitudinal dimension of the material stream).

According to the TOS, 1-D lots present the optimal sampling situation, preferring that 2-D and 3-D lots (e.g., industrial, geological or environmental strata, stacks, stockpiles, silos) should, where possible, be transformed to comply with a 1-D sampling situation [3,10]. In practice, this is often possible by locating another situation where the lot already is in transport.

Lot-dimensionality transformation constitutes one of the governing principles (GPs) of the TOS, described further below (Table 1). The reason for being this specific about lot dimensional-

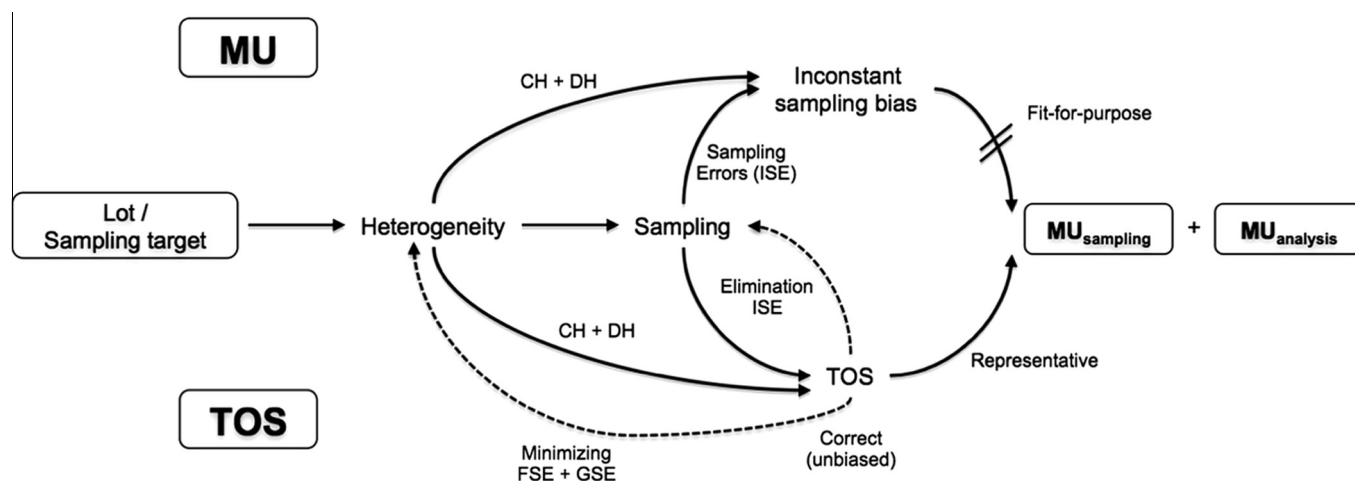


Fig. 2. Structural overview of similarities and differences between the TOS and MU pathways from lot to MU_{total} ($MU_{sampling}$ and $MU_{analysis}$). Fig. 2 also depicts the structure of the discourse in this article. If the effects caused by the inconstant sampling bias have not been properly dealt with, the ultimate MU_{total} cannot be termed representative, nor will it necessarily be fit-for-purpose (FFP). (FFP is defined in [12]: “The property of data produced by a measurement process that enables a user of the data to make technically correct decisions for a stated purpose”). It is evident that all sampling MU contributions logically must be included in the concept ‘measurement process’, so that the FFP criterion also critically depends on valid sampling-error treatment).

ity is the inherent (complex) heterogeneity of all naturally-occurring materials, which makes sampling far from a trivial materials handling issue. Proper understanding of the heterogeneity phenomenon, its influence on the sampling correctness and, most importantly, how heterogeneity can be counteracted in the sampling process require a certain level of knowledge. There is a need to be competent with respect to the TOS.

2.2. Heterogeneity

Below we present a sufficient minimum of the TOS tenets to allow full understanding and appreciation of deficiencies inherent in the current MU approaches. Before defining these concepts theoretically, Figs. 3 and 4 illustrate typical manifestations of the phenomenon of heterogeneity, which in practice has infinitely many manifestations.

For well-mixed materials [e.g., Fig. 4 [B], which are of identical composition to [A], or materials which apparently are ‘homogenous’ to the naked eye [D]], notions of simple sampling of homogenous materials have often been thought of as lending support to the statistical assumption of systematic, random variability components. (Powders will nearly always appear visually homogenous because of light-scattering effects, but the chemical composition of the individual particles may still span the complete range from identical to extremely different composition). But, homogenous materials comprise a minor proportion of materials with special

characteristics only, which can never justify generalization to all types of significantly heterogeneous materials. This is a critical issue underlying much of the following.

The theoretical analysis of the TOS of the phenomenon of heterogeneity leads to the recognition that the total material heterogeneity in a lot must be distinguished as two components [i.e. the constitutional heterogeneity (CH) and the distributional heterogeneity (DH), respectively], which are conceptually and mathematically defined in full only in the TOS.

CH describes the heterogeneity depending on the chemical and/or physical differences between individual “constituent units” in the lot (e.g., particles, grains, or kernels), which are generically termed “fragments” in a subtle, ingenious coverage also of the situation in which the sampling procedure accidentally or unavoidably fragments original particles. Note that each fragment (particle) can exhibit any analyte concentration in the range 0–100%. When a lot (L) is sampled, CH_L manifests itself in the form of a Fundamental Sampling Error (FSE) effect. CH_L increases when the compositional difference between fragments increases; CH_L can only be reduced by comminution (typically crushing). There will always be an influence from FSE in any sampling process; it can never be eliminated completely.

DH_L , the distributional heterogeneity complement, reflects the irregular spatial distribution of the analyte at scales between the entire lot and the sampling-tool volume (size of the increment, the correctly delineated and materialized unit of the lot. Incremental sampling implies that several increments are destined to become part of an aggregated composite sample. Sub-sampling is the opposite, divisive process, in which a sample is mass-reduced. It is crucial that this takes place in a representative fashion (SUO 10 in Table 1). DH_L is caused by the inherent tendency of particles to cluster and segregate locally (grouping) and more pervasively throughout the lot (segregation), or any combination thereof in a bewildering array of practical manifestations (e.g., Figs. 3 and 4). When sampling lots have a significant DH_L , there is a totally assured chance of non-representativity, when sampling is based on single-increment procedures (grab sampling). In the framework of the TOS, this is viewed as reflecting a specific Grouping and Segregation Error (GSE) in addition to FSE. DH_L can be counteracted by the process of mixing and/or by suitably-deployed, problem-dependent composite sampling with a sampling tool allowing a high number of increments [3,7,10,11,13]. Mixing is a very effective

Table 1

Axiomatic TOS synopsis – Six Governing Principles (GP1–6) and four Sampling Unit Operations (SUO 7–10)

1. GP	Fundamental Sampling Principle (FSP)
2. GP	Sampling Scale Invariance (SSI)
3. GP	Principle of Sampling Correctness (bias-free sampling) (PSC)
4. GP	Principle of Sampling Simplicity (primary sampling + mass reduction) (PSS)
5. GP	Lot Dimensionality Transformation (LDT)
6. GP	Lot Heterogeneity Characterization (0-D, 1-D) (LHC)
7. SUO	Composite Sampling
8. SUO	Mixing/blending
9. SUO	Comminution (crushing)
10. SUO	Representative Mass Reduction (representative sub-sampling)



Fig. 3. Heterogeneity has infinitely many manifestations (selected examples shown here only). Material/lot heterogeneity can be structured (e.g., as in a pegmatite intrusion (dike) [A] or in glacial till (soil) [E]), invisible to the naked eye (i.e. 'uniform materials') at many scales (e.g., wine grapes [B] or lightweight expanded clay pellets ('LECA') [D]), or the state of heterogeneity can be hidden from observation (see also Fig. 4), as in the proverbial 'big bag' case [F] and [C], which shows a process analytical technology (PAT) probe insertion in a pipeline (process sampling).

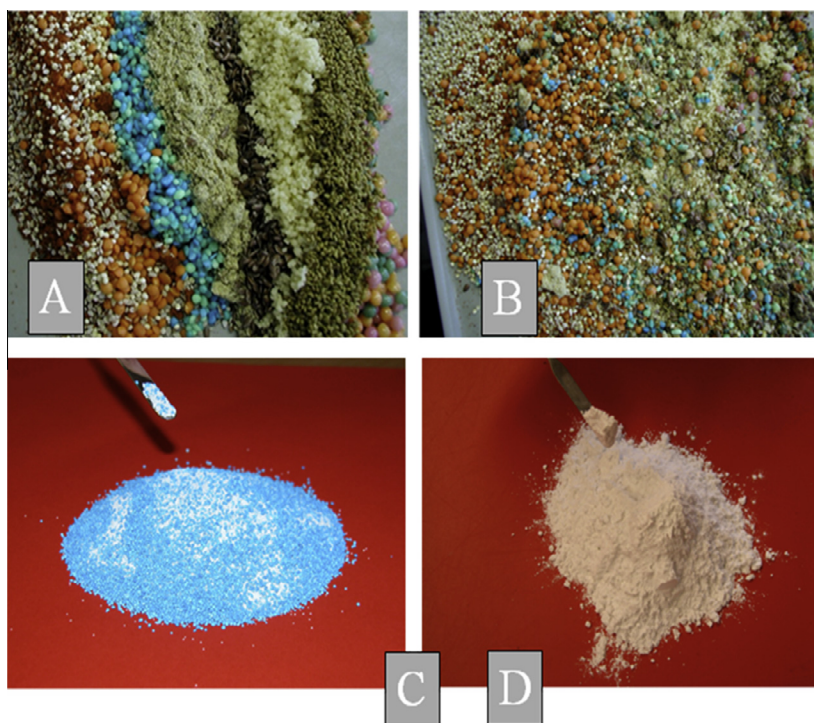


Fig. 4. Highly segregated heterogeneity (e.g., layering [A]) cannot be expected to follow any known statistical distribution, nor can the physical grain-size distribution shown in [C] (in which all grains with diameters above the average have been dyed blue, while all spherules with diameters below average remain white). [C] shows a strongly heterogeneous spatial distribution, which is purely physical, since all spherules are of identical composition. Even with this simplification, it is obvious that the statistical notion of modeling every heterogeneity manifestation within the concepts of systematic and random variability is too simple to cover the almost infinite variations of lot/material heterogeneity. In addition, overwhelmingly many lots most certainly do not consist of units (grains, particles, or other) of identical composition.

agent for reducing the spatial heterogeneity, but it is usually only applicable after the primary sampling stage (i.e. in the laboratory); it is very rarely possible to carry out forceful mixing of an entire

primary lot. Lots come in all forms, shapes and sizes spanning the gamut of at least 8–10 orders of magnitude (m/m) (i.e. from μg -aliquot precursors to MT industrial or natural-systems lots).

It is essential to understand, to acknowledge, and to act appropriately upon DH_L not being a permanent, fixed property of the lot. GSE effects cannot be reliably estimated, as the spatial heterogeneity is erratic in both space and time. In practice, DH_L is a transitory characteristic, because lots are, e.g., manipulated, transported, on-loaded and off-loaded. DH_L can be changed intentionally (reduced) by forceful mixing, but it can also be altered unintentionally (e.g., by materials handling or other agitations). It is one of the essential insights of the TOS that it is futile (even) to try to estimate a particular DH_L under an assumption of constancy. Instead, the TOS focuses on the necessary practical counteracting measures that will reduce GSE as much as possible (the goal is full elimination wherever possible) as an integral part of the sampling and sub-sampling process (but only in favorable instances will it be possible to eliminate GSE effects completely). It is thus more an act of faith that the notion of systematic *versus* random variability within populations (consisting of 'units' that are identical except with respect to the value of the measurand) can model all aspects of heterogeneous materials. Most of the materials in Figs. 3 and 4 demonstrably do not make up populations of identical units – instead the DH_L irregularity is overwhelming. Physical heterogeneity, especially spatial heterogeneity, is simply too irregular and erratic to be straightjacketed in this traditional statistical fashion. This issue was recently also debated in a more theoretical, scholarly exposé by Pitard and Francois-Bongarcon [14].

2.3. The sampling process

Perhaps paradoxically at first encounter, the TOS focus is not with 'the sample' but exclusively with the sampling process that produces the sample. Without specific qualification of the sampling process, it is not possible to determine whether or not a particular sample is representative. Loosely speaking of 'representative samples' without fully describing, fully understanding and fully documenting the lot provenance and the sampling process is but an exercise in futility (this includes massive confusions, such as 'more representative', or 'less representative'). Only a sampling process designed according to the rules of sampling correctness can produce representative samples. There is thus no declination of this attribute, a sampling process is representative or it is not representative.

The primary requirement in this context is sampling correctness, which means elimination of all bias-generating errors [termed 'incorrect sampling errors' (ISEs), see Fig. 2]. After this requirement has been achieved (by a correct sampling process), the main thrust in the TOS is to ensure an equal likelihood for all increments of the lot to be selected, without which all prospects for representativity are lost. This demand is known as the 'Fundamental Sampling Principle' (FSP).

According to the TOS, a sampling process is representative only when it is both accurate and precise [3]. A sampling process can be rated as accurate only if the average sampling error (m_e) [i.e. the difference between the analytical sample grade (a_s) and the average lot grade (a_L)] equals zero or results in a pre-determined, acceptably low value. (the TOS term 'analytical grade' is synonymous with 'measurand' in MU). Strictly within the TOS, the concept of "true (average) lot concentration" can still be used, although MU proponents will object since in MU both this term and "error" have been abolished. It is not the task of this article to resolve all theoretical, conceptual and terminological disagreements between the TOS and MU. (Below, we demonstrate how both TOS and MU can keep their respective terminologies amicably without adverse effects). Random effects (imprecision), caused by the FSE and the GSE [collectively termed the 'Correct Sampling Error' (CSE)], should subsequently be reduced as far as possible. A sampling process is only precise if the variance of the sampling error (e) is below a

predetermined, low threshold value. (The relative sampling error is defined as $e = (a_s - a_L)/a_L$, where a_L is the analytical grade of the lot and a_s the analytical grade of the sample. Sampling errors, and the notion of the 'true lot grade', a_L , play an essential role in theoretical developments of the TOS, from which all practical sampling procedures, among others, are derived, but they are not intended to be measured or estimated).

The relation between bias-generating errors, ISEs and CSEs, is depicted in Fig. 5, which sums up all elements recognized as potential contributors to the 'Total Sampling Error' (TSE). The term CSE signifies that these errors remain even when the sampling process is structurally correct, while the ISEs, if not eliminated, always cause a significant sampling bias. The effect from lingering ISEs can never be reliably estimated, as they will vary in magnitude for each re-estimation (such is the nature of material heterogeneity), causing the sampling bias to be inconstant. The sampling bias can consequently not be subjected to a conventional statistical correction for systematic effects (bias correction).

The TOS has analyzed the concept of heterogeneity in full, especially its manifestation in the sampling bias – and, by fully acknowledging these objective characteristics of all lots in science, technology and industry, the TOS reaches the conclusion that the ISEs must be eliminated and, for that, it describes all necessary countermeasures (see Fig. 2). There is therefore a logical demand in the TOS that all sampling processes must contain an active element of TSE reduction, preferably complete elimination, regarding ISE. This issue constitutes the primary conceptual discrepancy between TOS and MU.

2.4. Sampling errors – and their effects

The term 'error' in the TOS denotes a specific source that generates, or contributes to, the total MU. An important duality: while it is qualitatively essential to understand the origin and the circumstances influencing the source of specific sampling errors, it is only their manifestations (i.e. variances, or standard deviations) that can be estimated quantitatively.

The ISEs are three-fold: 'Increment Delimitation Error' (IDE), 'Increment Extraction Error' (IEE) and 'Increment Preparation Error' (IPE).

IDE relates to variations of the geometrical outline of the physically to-be-extracted increments, which can be avoided by stringently identical delineation of each increment.

IEE manifests itself, e.g., when particles that belong to the delineated increment do not end up here. This principle is also referred to as the "center-of-gravity rule", which states that particles with their center of gravity inside the delineated increment when intersected by the sample cutter edge(s) must end up in the final sample [3,10]. This requires that: no particles can bounce off adversely from the sampling-tool edges; no fine particles can be blown away or left behind before extraction; and, that particles outside the delineated increment should not be able to end up in the final sample (in order to avoid contamination).

IPE occurs when increments/samples are altered after extraction (which they should never be able to). In order to avoid effects, such as contamination, moisture absorption, evaporation, misidentification, loss of material or even fraud and sabotage, all samples require the utmost care in handling, correct sealing and storage. IPE is one sampling error, which can be completely controlled, but it critically depends upon strict, professional quality assurance/control of all processes, instrumentation and personal competence.

The TOS deliberately introduces the ISEs in order to signify that these errors, if not eliminated, always cause a significant sampling bias and are therefore the source of unpredictable, high sampling uncertainty.

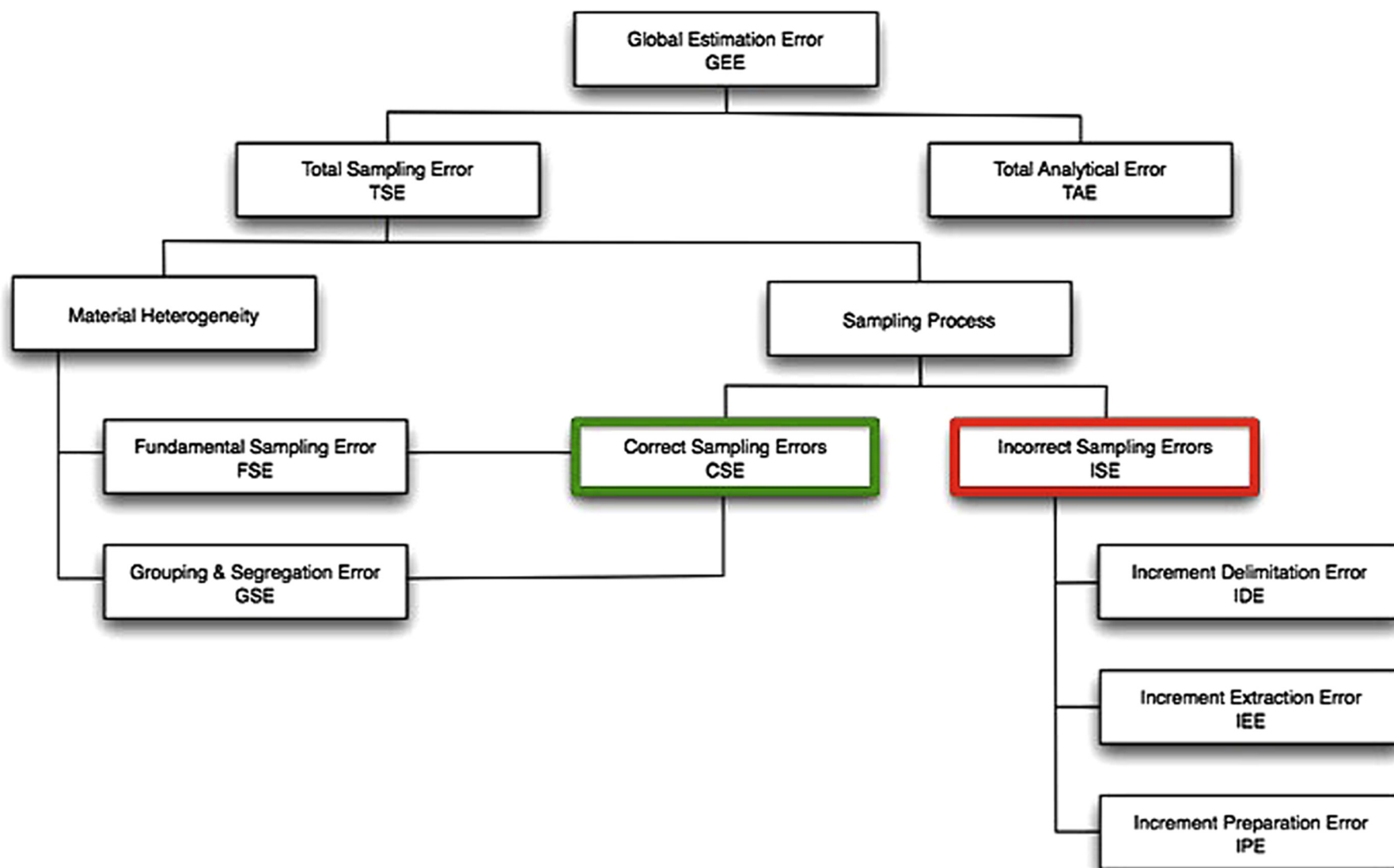


Fig. 5. Relationships of the five basic TOS sampling errors in stationary lot sampling. The incorrect sampling error effects originate because of faulty, ill-informed or wrongly-performed sampling processes. The correct sampling error effects originate because of interaction between the sampling processes with significantly heterogeneous materials – irrespective of whether or not the incorrect error effects have been properly eliminated. (Illustration source: [15], modified from [8]).

We do not deal in this article with the ‘Increment Weighting Error’ (IWE) and ‘Point Selection Error’ (PSE, which only affects process sampling) – two further bias-generating errors, which are not included in Fig. 5 – since established methods exist to eliminate the effects arising from these errors. This omission has no effect on the stated general objective of the present assessment. For completion, the PSE is important in any sampling target that cannot be treated as a 0-D lot, and not only in 1-D targets. The subject of 1-D sampling, process sampling, is dealt with fully in the dedicated TOS literature and there are introductions [3,5,10,16].

The TAE is identical to the total MU_{analysis} (see Section 3). TAE and TSE sum up to the ‘Global Estimation Error’ (GEE).

2.5. Sampling in practice

A recent unified approach for valid estimation of the GEE in the form of a new international standard, termed ‘DS 3077 Representative Sampling – HORIZONTAL standard’ [17], introduces the coefficient of variation (CV) as a measure of the effective total sampling variance, here called Relative Sampling Variability (RSV). For 1-D lots [i.e. sampling situations for which one dimension in time or space dominates compared to the other two dimensions of moving streams (width and thickness)], the variogram is a very useful tool to detect and to characterize process variations as a function of scale (i.e. distance between sampling units with different spacing, lags). For both process lots and equivalent stationary 1-D lots, the variogram expresses the 1-D heterogeneity along the effective singular dimension (see TOS literature and DS-3077 for details).

2.6. TOS summary

The systematic framework in the TOS of principles and operations for representative sampling enables us to evaluate the representativity of all types of sampling methods and equipment. Table 1 presents for the first time a complete axiomatic overview of the TOS in the form of six GPs and four sampling-unit operations (SUOs) (where no confusion can arise, these are sometimes also collectively referred to as the 10 SUOs of the TOS). The first six constitute GPs for designing and performing sampling processes, or modifying existing ones, while the last four constitute the only four practical procedures available for sampling purposes.

The TOS is comprehensive and complete, in the way that these 10 SUOs summarize all principles and practical procedures needed to ensure a correct (bias-free) and variance-reduced sampling along the complete lot-to-aliquot pathway, including all sample handling, mass reduction and sample-preparation steps in the analytical laboratory. More theoretical background to each SUO can be found in the TOS literature. First, it can only be rated representative when a given sampling procedure is correct (unbiased), and GSE and FSE have subsequently been minimized in order for the sampling procedure also to be sufficiently precise, as depicted in the TOS pathway in Fig. 2. Only on this basis can an uncompromised $MU_{\text{sampling} + \text{analysis}}$ estimate finally be assigned (Section 4).

3. MU pathway

The MU approach is discussed below following the pathway indicated in Fig. 2 (upper part). Deficiencies in GUM and the EURACHEM guide are pointed out. The main conceptual differences between MU and TOS are highlighted in explanations 1–7 at the end of this article. The EURACHEM guide receives special focus since users need methods for estimating the MU of the entire process from sampling, mass reduction and sample preparation (TOS) to the analytical measurement process (MU_{analysis}).

All sampling targets are very nearly always characterized by significant heterogeneity (i.e. deviating from the ideal homogeneity, which is defined in the EURACHEM guide [2] as “the degree to which a property or constituent is uniformly distributed throughout a quantity of material”). This definition is not comprehensive enough to deal with the many varieties and manifestations of heterogeneity, and far from concise enough to function as a guide for the user who wants not just to estimate the total MU ($MU_{\text{sampling} + \text{analysis}}$), but who also wants to reduce MU_{total} as much as possible (i.e. who wants the most realistic MU_{total} estimate).

The process of experimentally obtaining quantity values for a measurand is defined as ‘measurement’, requiring specified procedures and conditions. The MU “includes components arising from systematic effects, such as components associated with corrections [...], as well as the definitional uncertainty” [23,24]. [Sampling effects are not included in these systematic effects in MU (and there is no equivalent in sampling to the constant, systematic analytical bias), see also Explanation 4 below]. The concept of definitional uncertainty, signifying the “component of measurement uncertainty resulting from the finite amount of detail in the definition of a measurand”, is termed ‘intrinsic uncertainty’ in the ISO/IEC Guide 98-3:2008 [25]; this is the minimum uncertainty practically achievable in any measurement of a given measurand. GUM does not explicitly state procedures for estimating the definitional uncertainty. There is no analogue to such a concept in the TOS realm, which instead painstakingly analyses all error types and their effects in full, as laid out in detail below.

EURACHEM points out eight main sources that effect MU, of which the first two refer to sampling and sample preparation [26,27], stating that sampling uncertainty can be affected by heterogeneity, sampling strategy (e.g., systematic, random, or stratified random), physical state of material, effects of movement of bulk medium, temperature and pressure effects, effects of the sampling process on material composition, and transportation and preservation of samples. The uncertainty at the sample-preparation stage can be affected by physical factors, such as drying, milling, dissolution, extraction, loss of analyte, loss of fine particles or contamination [2,26,27]. In listing all such factors that can affect MU estimates, there is very little difference between TOS and MU, but a marked difference emerges regarding what to do about the first two factors.

Notable ‘gross errors’, such as involuntary mistakes (e.g., lack of knowledge, spillage, contamination, mixing of sample numbers), or deliberate faults are specifically excluded from these uncertainty estimates [2]. Many of this type of “mistake” should rather be understood as, and termed, ‘errors’ in analytical practice, as indeed they are in the TOS, where some are included in the IPEs and others make up parts of the ISEs (see discussion in Explanation 3).

For estimating the MU caused by sampling, the EURACHEM guide introduces two approaches – i) empirical and ii) modelling. These approaches can also be used in combination, if desired. The empirical method, also termed the ‘top-down’ approach, quantifies uncertainty by determining the effects caused by “factors such as the heterogeneity of the analyte in the sampling target and variations in the application of one or more sampling protocols” using “repeated sampling and analysis, under various conditions” [2]. The key MU focus remains to achieve a reliable estimate of the overall uncertainty focusing on reproducible estimates as acceptance criterion. This does not necessarily require knowledge about all individual uncertainty sources, since the MU approach focuses on the principal subdivision into random and systematic effects by the sampling or the analytical process, which, in MU, is manifested as sampling and analytical precision (random effects) as well as sampling and analytical bias (systematic effects).

The latter postulate of the analytical bias and an alleged analogous sampling bias constitutes the singular theoretically

(and practically) most important difference between TOS and MU, as outlined fully below, which cannot be overemphasized.

Random analytical effects are normally estimated by ‘duplicate measurements’, using conventional, effective statistical approaches; here, MU and TOS are in full agreement. Within MU, the analytical bias is to be estimated based on comparisons based on certified reference materials (CRMs, see Explanation 5). However, the situation may, or may not, be different when it is realized that the only relevant definition and configuration of ‘duplicate measurements’ must be realized by full ‘duplicate sampling’ from the primary sampling stage (always including the first sampling stage). This understanding forms the basis for the elaborate ‘replication experiment’ imperatives laid out in the new standard, DS 3077 Representative Sampling – HORIZONTAL standard [17]. Logically, any deviation from this demand [i.e. starting duplication (or replication) at any later stage than the primary sampling stage] is a breach of due diligence [17].

For estimating the sampling bias, the EURACHEM guide suggests the use of a reference sampling target (RST) as “an equivalent to reference material(s)” to estimate the bias, or alternatively to compile measurement results from “inter-organizational sampling comparison trials” (see Explanations 4 and 6).

Depending on which type of method is used (duplicates *versus* protocols: CTS *versus* SPT – EURACHEM [2], section 9.4), such an empirical approach allows four different experimental set-ups with which to estimate the total MU:

- (1) sampling plus analytical precision (duplicates);
- (2) sampling precision plus bias in-between different protocols, as well as analytical precision (protocols);
- (3) sampling precision plus bias in-between different samplers, as well as analytical precision and bias (CTS); and,
- (4) sampling precision plus bias in-between samplers and protocols, as well as analytical precision and bias (SPT).

These experimental plans are all analyzed by ANOVA. Below follows our critique of the notion of RST, which clashes with a full understanding of heterogeneity following the TOS. If the RST concept cannot stand up in its alleged role as a direct analogue (regarding sampling) to CRM (regarding analysis), EURACHEM options iii) and iv) are compromised.

The above assessment points out that the more complex EURACHEM experimental scenarios are not based on a fully comprehensive heterogeneity concept in relation to the many manifestations met with in science, technology and industry. They are specifically unable to cover the full range of challenges in sampling all types of heterogeneous materials and lots under the limited specifications offered. The key issue in the present critique concerns the lack of inclusion of the effect of ISEs, without which there can only result unnecessarily inflated MU estimates. Such MU estimates can therefore never be considered valid (or even fit-for-purpose if ISE effects remain in the definition of FFP).

We here advocate a much simpler, direct approach: the replication experiment (stationary lots) or variographic characterization (1-D lots), each of which works directly on the lot to be sampled and each of which captures the combined effect from the specific sampling procedure/material heterogeneity interaction without any of the excessive RST complications demonstrated above, see DS 3077 [17]. It is imperative to understand that each individual sampling procedure interacts with the same heterogeneous lot as all other alternative procedures and that each specific combination will produce different sampling-variability estimates and thus a different MU_{sampling} – and hence a different MU_{total} .

The EURACHEM guide refers, correctly, to variography for estimation of the combined MU from analysis and sampling in

the case of process sampling and process monitoring, in full agreement with the TOS (see Explanation 7).

Composite sampling is a fundamental issue on which TOS and MU agree substantially. In order to ‘cover’ lot heterogeneity appropriately, the logical approach is clearly by the use of composite sampling [i.e. by deploying an optimized number, Q , of correctly sampled increments covering the entire spatial geometry (volume) of the lot as well as possible within a set of given conditions]; it is manifestly not enough to specify only the number of increments to be used without this spatial coverage imperative. Note, however, that only the TOS enables the sampling operator to establish correctness (un-biasedness), which is absolutely not an automatic attribute of any sampling equipment or procedure by itself (design, operation and maintenance of procedures and equipment must be so that ISE effects are eliminated – not a trivial task, but a necessary task nevertheless). These conditions are often not fully understood. The TOS is the only framework that furnishes ways and means with which to optimize Q in relation to the empirical heterogeneity met with, either via replicate experiments or by variographics [see, e.g., DS-3077 [17] and references therein].

4. TOS – the missing link: a call for integration

The above evaluation of GUM and the EURACHEM guide shows that MU is not fully comprehensive, and is not a universal, guaranteed approach to estimate an uncompromised total MU from sampling. Sixty years of theoretical development and application of TOS practice has shown that sampling, sample handling and sample-preparation processes are associated with significantly larger uncertainty components than analysis (measurement) itself, multiplying MU_{analysis} typically many times over: in the range $10\text{--}50 \times \text{TAE}$, obviously very much dependent upon the specific lot heterogeneity in question.

While GUM focuses on $MU_{\text{analytical}}$ only, the EURACHEM guide does point out some of the potential sampling-uncertainty sources, but then leaves samplers incomplete and without the necessary means to take appropriate actions regarding sampling errors (some recognized, others neglected in MU, see below). The present critique has indicated that only the TOS specifies which types of errors can, and should, be eliminated (ISE) and which cannot be eliminated, but should instead be minimized (CSE), and, crucially, how. It is manifestly impossible to acquire sufficient conceptual understanding (CH/DH) and practical sampling competence with respect to these critical success factors for representative sampling from the MU literature in its present form. In the conceptual framework of MU, ISEs of the TOS are non-existent, and the GSE is only considered to an incomplete extent, leaving the TAE and the FSE as *de facto* the only main sources of MU.

Furthermore, in EURACHEM’s four empirical approaches to MU, the scope of the MU estimate depends on the method applied. Only the sampling proficiency test (SPT) approach considers analytical precision and bias, and sampling precision and bias, albeit in abridged form only. Otherwise, the sampling bias is considered to only a severely limited extent. It is tacitly assumed, but incorrectly so, that a sampling bias can be likened to a systematic effect in the standard statistical understanding. However, the physical nature of the sampling bias is most emphatically not of this simplistic nature – the main feature of the sampling bias is its very violation of constancy. For these reasons, the only scientifically acceptable way to deal with any and all sampling bias is to eliminate it.

Thus, then, is where the major distinction between TOS and MU becomes clear: the TOS notion that a sampling bias is a reflection of the ISE effects interacting with a specific heterogeneity *versus* MU’s notion of a statistical bias resulting from systematic effects attributable only to in-between protocols (SPT) and/or in-between

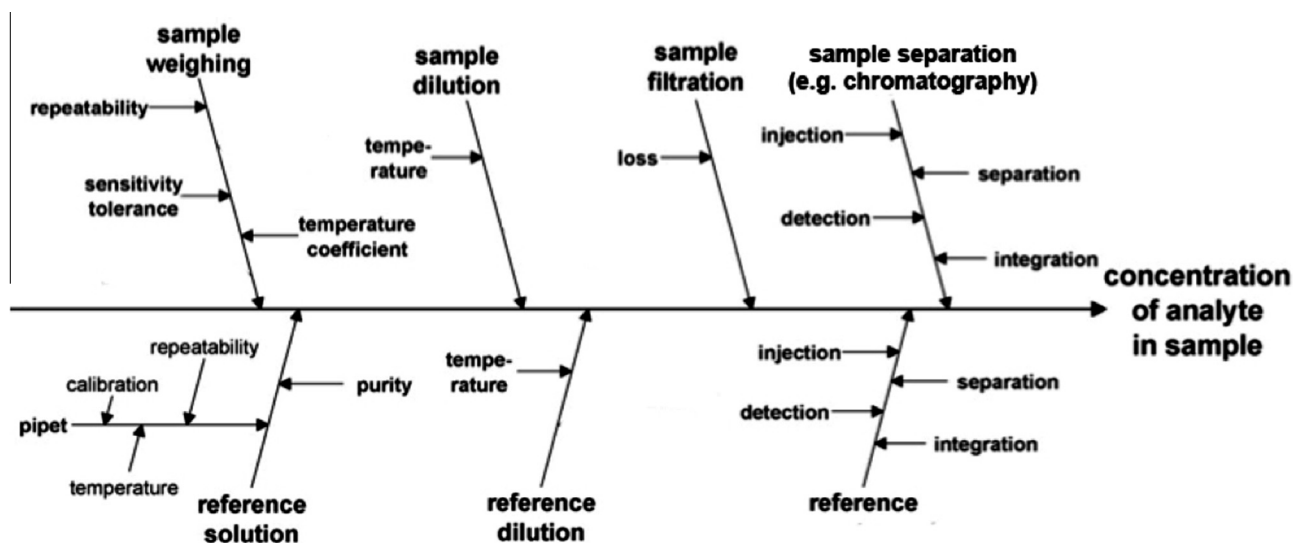


Fig. 6. Exemplar MU fishbone flow-path diagram with standard MU measurement uncertainty sources. Source [36], redrawn and simplified by the present authors.

samplers (CTS). Since the empirical MU approach is a top-down approach, dependent upon the assumed framework of random and constant systematic effects, the individual uncertainty sources, such as GSE and ISE, are not subject to separate identification, concern, or estimation, or to the appropriate action (elimination or reduction).

The different realms of MU versus TOS are in serious need of clarification and reconciliation. The issue is not helped by the different scopes of GUM and the EURACHEM guide. However, these issues, while complex, can be distinguished naturally, and fully comprehended, following the TOS domain and terminology in the following simple framework:

- (1) GUM: MU considers (only) TAE
- (2) EURACHEM – empirical approach (SPT): MU considers (only) TAE + FSE + GSE
- (3) EURACHEM – modelling approach: MU considers (only) TAE + FSE

GUM focuses overtly on uncertainties related to analytical measurement (i.e. weighing, preparation, dilution, filtration, handling and similar issues for reference materials), disregarding key elements governing all prior sampling and laboratory sub-sampling, among other issues, and their uncertainty contributions. The scope in the EURACHEM guide varies depending on the applied approach. The empirical approach, based on repeated sampling and analysis, includes FSE and GSE, since both of these will always be reflected in repeated sampling and analysis. {A subtle point here is that, whenever “re-sampling” is involved, it must always be replicated from the primary stage (see e.g., [17] and also further below). This is a critical criterion, in order for the final augmented MU estimate to be comprehensive and valid}. However, even the most comprehensive MU approaches focus only on the bias originating between sampling protocols (SPT) or the bias caused by different personnel performing the sampling (collaborative trial in sampling).

On this basis, it is clear that none of the stated MU approaches is able to estimate the full total GEE ($GEE = MU_{total}$), and can do so only partly under given, restricted conditions. A complete, and therefore optimal, MU approach must be defined in the following way:

- (4) Representative TOS approach: $GEE = TAE + TSE$, where $TSE = CSE + ISE$
- (5) Complete MU approach: $MU_{total} = TSE_{sampling} + MU_{analytical}$

Regarding (5), it falls to the TOS to take responsibility for the estimate of $MU_{sampling}$ at all stages along the lot-to-aliquot pathway. The necessary understanding and competence required is outlined by the minimum TOS framework presented here (and in [17]). It is specifically not enough to rely on claims of original equipment manufacturers (OEMs) regarding equipment and products. Many studies, audits and extensive practical consulting experiences have shown unambiguously that many OEMs producing and describing “representative sampling devices” and *ditto* “mass reduction equipment” are in fact marketing incorrectly designed products, often causing severe sampling bias (general TOS literature) and/or unnecessarily inflated total sampling variances. Sufficient TOS competence cannot automatically be taken for granted.

We here draw the logical conclusion to the above analysis and assessment. We call for integration of the TOS with the MU approach, easily illustrated based on the widely-used fishbone flow-path diagram. Fig. 6 shows a standard fishbone diagram depicting the standard complement of MU sources of an exemplary analytical measurement process (this may be even more comprehensive, without influencing the present conclusions). The uncertainty sources connected to sampling (i.e. both sample extraction and those preparation stages also involving sampling are completely disregarded). It is simply assumed that the analytical sample, which ends up as the test portion, has been extracted and mass reduced in a representative fashion. If this assumption does not hold, it is a sure guarantee that the appropriate TOS approaches have not been involved, and that the uncertainty estimate of the analyte concentration is invalid and of little value; it will inherently and unavoidably be too small by an unknown, but significant, factor, so it will also be invalid as a proper fit-for-purpose MU estimate.

In order to prevent structural underestimation of the full complement of active uncertainty sources, it is necessary to integrate the effects related to all sampling stages involved with this standard $MU_{analysis}$ scheme. This can be done in a perfectly seamless fashion (i.e. there need not be any changes regarding $MU_{analysis}$, while the framework surrounding $MU_{sampling}$ will be supplied by the TOS). Thus, Fig. 7 outlines all uncertainty sources related to sampling as a new main branch added to this diagram. The sampling branch of the TOS should be implemented in every MU fishbone diagram, left justified, signifying that all sampling uncertainty contributions must be dealt with before any of the traditional MU issues. Note (compare Fig. 1) that the sampling

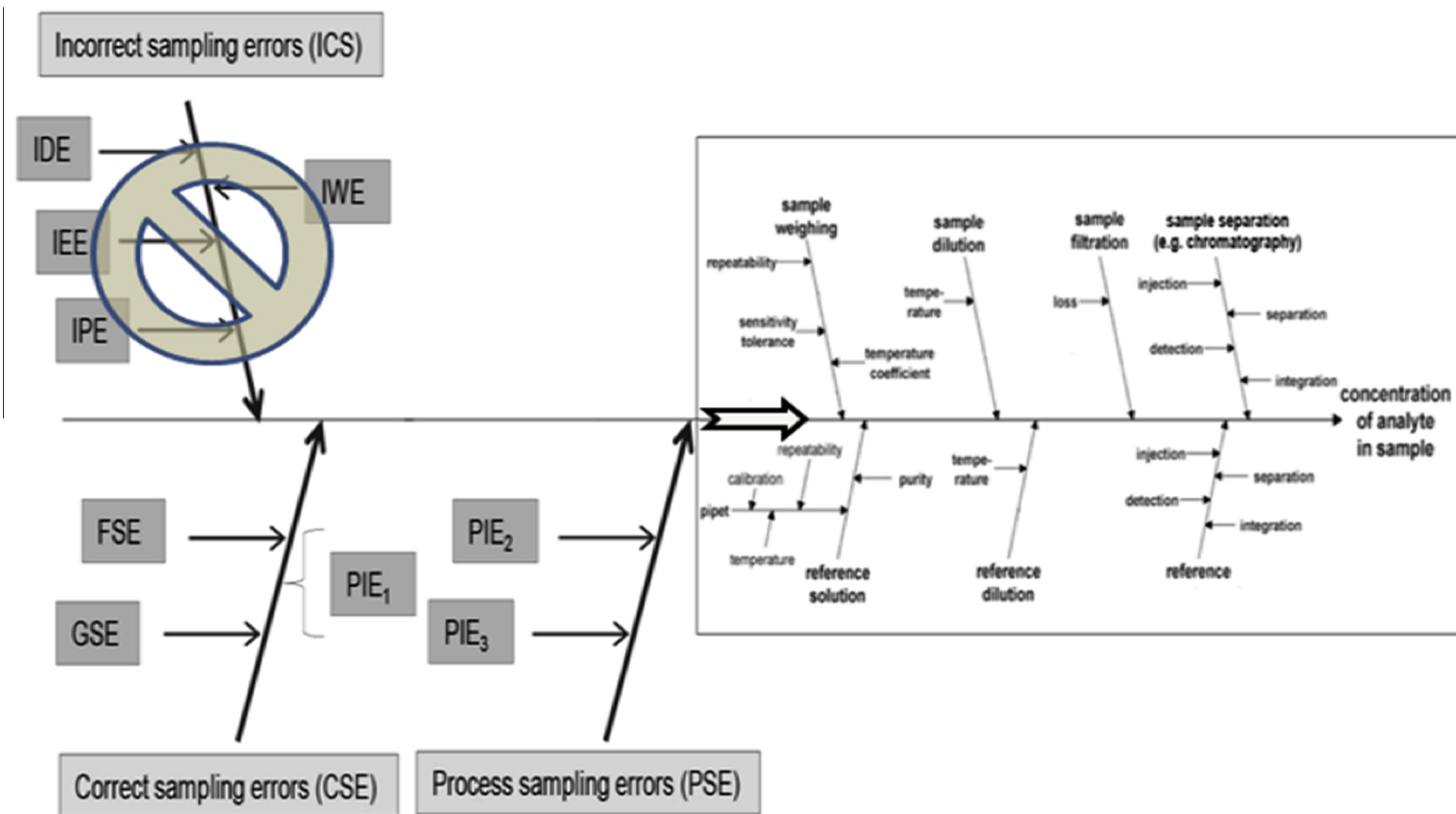


Fig. 7. Induction of all principal sampling uncertainty sources of the TOS in an augmented MU framework. The standard MU_{analysis} fishbone diagram is shown on the right, to which TOS is charged with delivering a representative analytical aliquot (arrow). This diagram illustrates the proposed TOS/MU integration in full, focusing on the imperative to eliminate incorrect sampling errors (ISEs).

responsibility of the TOS is imperative: the TOS is the agent responsible for delivering the valid, representative analytical aliquot for the subsequent $\text{MU}_{\text{analysis}}$ estimation.

The TOS stipulates that all ISEs must be eliminated (sampling correctness), followed by reduction of the remaining CSEs (and PSEs, if relevant) (sampling precision), until compliance with representativity and/or until a fit-for-purpose criterion. Note that the Point Integration Error PIE_1 is identical to (FSE + GSE), while PIE_2 (process trend error) and PIE_3 (process cyclicity error) are conveniently and easily dealt with within the context of the TOS paradigm for process sampling (see relevant process TOS literature). For the present purpose, there is no need to detail process sampling further, other than to note that it is already located in its logical place in the TOS fishbone branch.

5. Conclusion

A critical assessment of GUM and the EURACHEM guide shows that not all influential uncertainty sources are considered as to their full MU impacts. In particular, effects caused by ISEs are insufficiently defined and integrated. While GUM exclusively focuses on estimating the analytical MU, the EURACHEM guide indicates and incorporates some error sources related to sampling, but detailed analysis of the scope also here revealed several deficiencies compared to the full sampling-error framework of the TOS. While the EURACHEM guide acknowledges the existence of the CSEs, it stays with the assumption that all other sampling-uncertainty-error sources have been eliminated by other parties – which gives no help to the sampler/analyst. By excluding both the concept of, and the risk incurred by, the inconstant sampling bias, the sampler/analyst may well not even beware of the risk that the effective MU estimate will be principally different each time that it is re-estimated. The user is left without the crucial understanding that ISE effects will unavoidably result in uncontrolled and unquantifiable, inflated MU_{total} estimates.

Only the TOS offers complete theoretical and practical understanding of all key features related to heterogeneity and full practical insight into the intricacies of the sampling process when confronting the gamut of heterogeneity manifestations. Closing this gap between TOS and MU necessitates a certain minimum TOS competence, and confidence, that all sampling processes can indeed become correct (sampling free of bias), opening up for them also to become representative, or fit-for-purpose, where appropriately defined. This minimum competency has recently been outlined in a new international standard, DS 3077 [17], the history of which has been outlined [37]. In order to derive a valid estimate of the complete uncertainty for any measurement procedure (sampling and analysis), all ISEs and CSEs, as well as the TAE ($\text{MU}_{\text{analysis}}$), must have their proper place in the suggested augmented $\text{MU}_{\text{sampling} + \text{analysis}}$ context (Fig. 7). This opens the way to a unified sampling-and-analysis responsibility.

A detailed analysis of MU and formulation of the requirements for a universally optimal MU concept outlined the critical deficiencies in MU and pointed out that the TOS can simply be inducted as an essential first part in the complete measurement-process framework, taking charge and responsibility of all sampling issues at all scales (i.e. along the entire lot-to-aliquot process).

We here call for a constructive integration between TOS and MU, allowing reconciliation of these two frameworks that all too long have been considered only antagonistically.

6. Postscript: terminology issues

One could perhaps conceive of a potential terminology debacle in the wake of the present proposal. For one thing, MU denounces

with extreme prejudice the notion of “error” and “true value” (as in “sampling error” and “true average lot concentration, a_L ”), among others, and has replaced this tradition with a well worked out, highly systematic MU conceptual alternative terminology, as codified in VIM3 [23,24]. While this is a relevant development within analysis, when addressing sampling in the full understanding of heterogeneity, this becomes a severely impractical straight-jacket and a battleground of immense futility. As it turns out, it is quite unnecessary. Both TOS and MU can simply be left with their separate terminologies and can fulfil their complementary roles unaffected. From the comprehensive theoretical analysis of the TOS of the phenomena of heterogeneity, sampling procedures, and sampling equipment is derived the superior practical command of practical representative sampling, allowing all types of lots to be sampled, not just those associated with various restricted understandings of heterogeneity with a concomitant desire to view all heterogeneity issues in the statistical notion of systematic effects and stochastic variability only.

A scientific concept and terminology skirmish is also uninteresting in view of the separate histories and the complementary practical roles of TOS and MU. Neither framework can win such a battle in view of their hitherto individual histories, achievements and their present status. The only constructive way forward lies with the proposed integration and reconciliation.

7. Explanation 1. Heterogeneity

Heterogeneity, introduced above, is the prime characterization of all naturally-occurring materials, including industrial lots, intermediate materials and products, processed and manufactured materials, and all materials in the natural world. Rocks could serve as an example of significantly heterogeneous materials in the natural world (also mineralizations, polluted sediments, toxic wastes, mineral-processing streams, commodity raw materials) – the range of examples from all of science, technology and industry is legion. Moreover, heterogeneity manifests itself *at all scales* related to sampling from residing inside grains, contributing to CH_L between grains, occurring at meso-to-lot scales as ‘grouped’ fragment clusters and as segregation, from incipient to pervasive. Thus, heterogeneity manifests itself everywhere in the scale hierarchy from grain to lot, and the issue rather concerns to what degree all substances are heterogeneous (see below for a very few, marginal exceptions).

It is much more than a quibble, to point out that heterogeneity should be defined as the degree to which a property or a constituent deviates from an assumed uniform distribution throughout a quantity of material, instead of the degree to which it conforms to an unrealistic ideal concept of random distribution. It is counter-productive to keep to the ideal notion of a uniform distribution, because such is never the case for the very many, very different types of materials and lots that are to be sampled.

Uncritically taking on the notion of a random distribution, which can then be considered fully with traditional statistical tools, is a very dangerous endeavor. An example of quantitative analysis of genetically-modified organisms (GMOs) provides a poignant case. Esbensen et al. [18,19] and Minkinen et al. [20] outlined in detail the consequences of carrying over the notion of random distribution of the “property of interest”, in this case quantitation of GMO in soy-kernel lots, and exposed many negative ramifications of such an attitude, which turned out to clash rather spectacularly with reality; above all, it was proved that conventional statistical estimates could be seriously compromised and frequently off by factors of 2–5. Thy et al. [21] demonstrated similar destructive effects regarding biomass-energy assessments, also originating with unsubstantiated random-distribution assump-

tions; consequences included published ash compositions, that could be proved to differ from true compositions by as much as factors of 2–3 for many major oxides (even the sign could be wrong in certain mass-balance calculations). Minkinen and Esbensen [22] detailed the reasons behind such faulty assumptions and, in particular, showed the very serious consequences of using grab sampling instead of composite sampling.

The MU definition of heterogeneity is incomplete in that it specifically only addresses one of the two aspects of heterogeneity, spatial heterogeneity (DH). This is unspecific (e.g., regarding analytes that may reside wholly or partly inside certain types of particles but not in others), and particles may obviously have widely different concentrations of a dispersed particular property. Particles may also be broken up during sampling (or they may not), partly or fully ‘liberating’ the “property of interest”, as a function of the sampling process. The compositional heterogeneity concept is not defined in MU.

The situation is somewhat more relaxed concerning the definition of “practically homogeneous materials”, termed “uniform materials”, which are defined as materials with a “repeated sampling reproducibility lower than 1%”. Claims have also been made that “small items” (presumably meaning “small lots”) are also not in obvious need of elaborate sampling instructions. However, such materials and cases only occur naturally in but the rarest of instances (e.g., exceptions are gases, well-mixed solutions, and manufactured pure or ultra-pure materials). However, the most important characteristic from such cases is that generalizations based on them with respect to sampling can never be valid for the gamut of all other types of materials and lots. It is by far the simplest always to treat all types of lots, including such marginal cases, all materials and sampling targets as examples of materials displaying significant heterogeneity, thus opening up for a universal sampling practice for all materials irrespective of their inherent degree of heterogeneity: All lots should be treated in identical fashion (i.e. as significantly heterogeneous lots). By way of completion, depending on the analytical viewpoint, a lot can simultaneously be both extremely homogenous – and extremely heterogeneous: while a lot consisting of an ‘ultrapure’ material can be regarded as homogeneous for almost all most practical purposes (e.g., if the concentration of the analyte is, say, 99.9999% (or higher), the lot can also be viewed as extremely heterogeneous – if the analytical focus is on impurities at ultra-low levels, say of the order of pg/g, ng/g (or below), which, by necessity, must be extremely irregularly distributed.

8. Explanation 2. Intrinsic uncertainty

The TOS defines the Global Estimation Error (GEE) as the sum of the Total Analytical Error (TAE) plus the Total Sampling Error (TSE). TAE is identical to the total analytical MU, MU_{analysis} . TOS defines a “minimum uncertainty”, called the Minimum Possible Error (MPE), which is concerned with the minimum sampling plus analysis uncertainty in practice, and which is related, at the very least, to the Fundamental Sampling Error (FSE), treated in more depth below, to which is added MU_{analysis} proper. The MU term “definitional uncertainty” relates to the analytical measurement process, but there is no equivalent needed for the sampling process because of the theoretical completeness of the TOS.

9. Explanation 3. Gross errors versus incorrect sampling errors

The TOS considers the first type of ‘gross error’ as part of the ‘Incorrect Preparation Error (IPE)’. This definition also allows inclusion of the effects of ‘gross errors’ in the overall Global Estimation Error (GEE), if they can be quantified, (see Fig. 3), as they most

certainly will contribute towards an inflated GEE. However, the effects from the IPE do not have to follow a tractable statistical distribution – the TOS specifically describes why this can never be. A key deficiency in the MU is that sampling errors, especially those that the TOS declares as of ‘major influence’, are excluded from attention simply by declaring these as ‘gross errors, which are assumed to have been taken care of before the MU. It has even been suggested to include the ISE in the definitional uncertainty while accepting the effects of FSE + GSE in the MU.

One of the major issues of dissent between the TOS and the MU concerns this twilight status and deliberate neglect of the incorrect sampling error (ISE) (the second type of ‘gross error’ in the MU). In the TOS, this would be unthinkable, if for no other reason than these dominate the total uncertainty budgets if not heeded properly, but also because they are indeed, and manifestly, subject to directed action: the TOS actively reduces, and seeks to completely eliminate, the effects from these critically important errors as part of a reconciled TOS/MU.

10. Explanation 4. Systematic measurement error versus sampling bias

Systematic error effects caused by sampling heterogeneous lots, termed ‘sampling bias’ in the TOS, are not constant, and therefore not “predictable”. A specified sampling procedure interacting with a given heterogeneous material will, if replicated, never result in an identical bias estimate precisely because of the nature of the material heterogeneity. Lot heterogeneity is a complex spatial and compositional feature characterizing the lot volume at all scales above the sampling-tool size, and it is transient (i.e. varying if/when the lot is manipulated in connection with sampling, or resulting from transportation). Sampling procedures that compromise the ‘correct sampling imperative’ (GP 3 in Table 1, fully defined in the TOS) will by necessity lead to effects, which, in the TOS, are attributed to the incorrect sampling errors (ISEs). Even when replicating a sampling procedure in a “100% identical fashion”, the resulting alternative analytical results will *per force* come out as different measurand values (concentrations) because of the pervasive irregular nature of heterogeneous materials. In other words, when replicating a sampling procedure, it is another primary increment of the heterogeneous target lot, which is extracted and subjected to the sampling analysis pathway – and for which, consequently, the analytical results, a_s , must be different.

While this difference at times may be negligible or small (small lots and/or uniform materials), and therefore perhaps ultimately only constitute an acceptable MU contribution, it may equally well deviate to a significant degree, depending on both the nature/magnitude of the lot heterogeneity in question and the sampling procedure used, either way leading to an unacceptable, unnecessarily inflated MU_{sampling} . The crucial issue is that it is never known *a priori* which of these alternative situations will be encountered, where or when. The only rational scientific attitude in view of such fundamentally incomplete knowledge is to act as if the adverse effect is always present and significant.

Above all, this principal uncertainty can never be used as justification for deviating from the strict rules of the TOS formulated to guarantee representativeness. Unfortunately, a varying, ‘inconstant sampling bias’ cannot be compensated for by any known means (e.g., data, analytical, statistical, equipment, or procedure), which all presumes a ‘predictable’ (i.e. constant) bias. The TOS allows all samplers the easy and full understanding that unrecognized, or uncontrolled, the ISEs create the inconstant sampling bias, so there is only one conclusion: ISEs must be eliminated from the specific sampling process involved. This solution is both logical and practically achievable. Occasionally, a fit-for-purpose (FFP) version

of this imperative may suffice, provided that the TSE + TAE effects are always subjected to proper estimation before acceptance or rejection of a FFP criterion.

11. Explanation 5. Certified reference materials (CRMs)

The approach to estimating the analytical bias is designed to work in the analytical laboratory, where every systematic effect can, in principle, be brought under control. However, it is only fair to point out that this is critically dependent upon 'TOS-correct' sub-sampling from batches of certified reference materials (CRMs), as received from relevant suppliers. But laboratory sub-sampling representativity is very often assumed without proper validation, despite sub-sampling always critically depending upon the "effective heterogeneity" of the CRM sachets supplied (e.g., containers, or vials). Such sachets are lots in their own right, albeit small, the only difference is in scale. As such, the critical issue here is, as always, how sampling is performed, in this case how the relevant sub-sampling is performed [8]. It is fully possible to conceive of unnecessary Total Sampling Errors (TSEs) even at this ultimate sub-sampling stage, as these have been demonstrated on numerous occasions. This issue serves well to underline that all TSEs are a result of a sampling procedure interacting with a heterogeneous lot at absolutely all scales.

This issue is far from trivial, as witnessed by numerous discussions and focused CRM heterogeneity studies (e.g., in journals, such as *Geostandards and Geoanalysis Research* (GRR), *Analytical Chemistry*, *Analytica Chimica Acta*, and *The Analyst*). This issue sometimes also includes certain aspects of the efficiency of dissolution of whole-sample materials [28,29]. The issue was well summarized, but acknowledged to be far from solved [30]. This issue also has a critical bearing on the MU issue regarding "sampling targets" (see further below).

To the degree that a CRM sachet is heterogeneous at the scale of a few test portion masses, say 5–15 or so, there is a very real danger of sampling errors also affecting even this ultimate sampling step producing the analytical CRM aliquot. This is why many calls have been made to supply CRMs with an effective "sampling constant" specifying a minimum sampling mass (sometimes augmented by a demand for a representative grain-size distribution documentation) [31,32] and further references herein. Many spectroscopic and image-analysis methods only get information from a relatively shallow surface layer of the final test portion, depending on the operative wavelength(s). In such cases, the reduced volume from which the information is obtained is the effective test portion and the possibility for significant sampling errors has to be considered even at this final measurement step also for these types of analysis. A basic introduction to these issues was given by Ramsey [33] {see also [31,34]}

12. Explanation 6. RST (reference sampling target), SPT (sampling proficiency test), and CTS (collaborative trial in sampling)

Analysis of Variance (ANOVA) is a statistical approach for variance decomposition proportioned along a set of experimental factors in the experimental design employed. The defining issue is that each factor is controllable by the fashion that the experimenter is able to set the specific levels desired in a design of experiment (DOE), or by a random factor. Different sampling procedures or sampling plans can, with a stretch, be codified as 'levels' on a sampling-mode factor, or different between-sample-distances may also be viewed as 'levels'. However, there would appear to be little or no possibility of 'different degrees of heterogeneity' (and its interactions with alternative sampling procedures) to be

similarly codified on an experimental factor, at least not without a truly staggering amount of work. More importantly for significantly heterogeneous lots in the real world, it is unrealistic to contemplate that an RST can ever be constructed precisely because of the compositionally complex and varying spatial heterogeneity involved (amongst others, even laying up the RST would, e.g., be subject to inconstant segregation effects). Above all, the RST approach is extraordinarily difficult and prohibitively laborious because one would first have to try to estimate the effective heterogeneity of the target lot reliably (indeed this itself must involve extensive sampling, not yet documented representatively) and then to try to construct a reference-sampling-target lot with identical heterogeneity characteristics, from which to try to obtain insight as to how to best to sample the original target. There appears to be no way such an endeavor can ever come close to simulating the entire target lot, without taking this apart *in toto*. This circular reasoning impasse has severe implications for the application potential of SPT and CTS. This approach is impossibly far away from general representative sampling, which, by way of contrast, can be easily accomplished based on the TOS, Table 1. The RST suggestion appears more to be offered because this makes it possible still to apply ANOVA to decompose the variance proportions commensurate with the DOE. Indeed, Ramsey [33] states that: "RST is still at the 'proof-of-concept' stage and yet not widely available". Against this, there exists extensive experience, substantiated in a formidable amount of TOS literature, attesting that real-world heterogeneity is far more complex than can ever be described with standard statistical techniques, notwithstanding their well-proven applicability in many other areas.

13. Explanation 7. Variographics (TOS)

The variogram is a powerful tool with which to characterize 1-D variations and which benefits from the inherent auto-correlation between units (increments, or single samples) sampled with different 'between pairs-of-samples' distances, termed 'lags'. Variography is particularly relevant for process sampling (or, equivalently, stationary 1-D lot sampling), both instances referring to lot configurations for which one elongated dimension in time or space dominates completely, because the other two dimensions are eliminated by the TOS stipulation that all increments (samples) must cover both these dimensions completely – hence the rigid demand in the TOS only to use correct increment delineation and extraction. The extensive approach of the TOS to both stationary and dynamic 1-D sampling addresses, e.g., moving streams of matter on conveyer belts or in pipelines, units transported as truckloads, railroad cars or tank vehicles, and manufactured or produced units, such as containers, vessels, or bags. Depending on their intrinsic heterogeneity characteristics, such streams are characterized by various degrees of 1-D auto-correlation, as manifested by the variogram. There exist numerous, in-depth descriptions, illustrations and very many case histories involving variograms in the TOS literature [3–5,10,35].

A comprehensive description of sampling streams of extremely irregularly distributed trace concentrations, including a thorough exposé of the versatility of variographic characterization as a general approach for designing "fit-for-purpose" sampling plans, commensurate with the empirical lot heterogeneity, can be found [18–20] dealing with batches of genetically-manufactured organisms (GMOs) originating as cargo shipments into Europe from international ports of origin. While specifically addressing feed, this methodological treatment is rather a complete exemplar, which can be applied to all similar 1-D lot heterogeneities; sampling of ultra-low-abundance precious-metal phases in minerals processing and polluting solids in natural stream water are but two examples.

References

- [1] GUM, Evaluation of measurement data – guide to the expression of uncertainty in measurement, JCGM 100 (2008).
- [2] EURACHEM, CITAC, 2007. Measurement uncertainty arising from sampling: a guide to methods and approaches. Eurachem, first ed. In: M.H. Ramsey (University of Sussex, UK), S.L.R. Ellison (LGC, UK) (Eds.), EUROLAB, Nordtest, UK RSC Analytical Methods Committee.
- [3] P.M. Gy, Sampling for Analytical Purposes, John Wiley & Sons Ltd, Chichester, 1998.
- [4] P.M. Gy, Proceedings of first world conference on sampling and blending, Special Issue Chemometr. Intell. Lab. Syst. 74 (2004) 7–70.
- [5] K.H. Esbensen et al., Representative process sampling – in practice: variographic analysis and estimation of total sampling errors (TSE), Proceedings 5th winter symposium of chemometrics (WSC-5), Samara 2006, Intell. Lab. Syst. 88 (1) (2006) 41–49.
- [6] K.H. Esbensen, P. Minkkinen, P. (Eds.), Special issue: 50 years of Pierre Gy's theory of sampling: proceedings: first world conference on sampling and blending (WCSB1). Tutorials on sampling: theory and practice. Chemometr. Intell. Lab. Syst. 74 (1) (2004) 1–236.
- [7] K.H. Esbensen, L.J. Petersen, Representative sampling, data quality, validation – a necessary trinity in chemometrics, in: S. Brown, R. Tauler, R. Walczak (Eds.), Comprehensive Chemometrics, vol. 4, Elsevier, Oxford, 2010, pp. 1–20.
- [8] L. Petersen, C.K. Dahl, K.H. Esbensen, Representative mass reduction: a critical survey of techniques and hardware, Chemometr. Intell. Lab. Syst. 2004 (74) (2004) 95–114.
- [9] L. Petersen et al., Representative sampling for reliable data analysis. Theory of sampling, Chemometr. Intell. Lab. Syst. 77 (1–2) (2005) 261–277.
- [10] F.F. Pitard, Pierre Gy's Sampling Theory and Sampling Practice, second ed., CRC Press, Boca Raton, 1993.
- [11] F.F. Pitard, Pierre Gy's Theory of Sampling and C.O. Ingamell's Poisson Process Approach. Pathways to representative sampling and appropriate industrial standards. (Dr. Techn. thesis), Aalborg University, Campus Esbjerg, Denmark, 2009. ISBN: 978-87-7606-032-9.
- [12] M. Thompson, M.H. Ramsey, Quality concepts and practices applied to sampling an exploratory study, Analyst 120 (1995) 261–270.
- [13] P.L. Smith, A primer for sampling solids, liquids and gases – based on the seven sampling errors of Pierre Gy, ASA-SIAM, USA, 2001.
- [14] F.F. Pitard, D. Francois-Bongarcon, Proceedings world conference on sampling and blending 5 (WCSB). Demystifying the Fundamental Sampling Error and the Grouping and Segregation Error for practitioners, Gecamin Publishing, Santiago, Chile, 2011, pp. 39–56.
- [15] C. Wagner, K.H. Esbensen, A critical review of sampling standards for solid biofuels – missing contributions from the Theory of Sampling (TOS), Renew. Sustain. Energy Rev. 16 (2011) 504–517.
- [16] L. Petersen, K.H. Esbensen, Representative process sampling for reliable data analysis – a tutorial, J. Chemom. 2005 (19) (2005) 625–647.
- [17] DS-3077, 2013. Representative sampling – HORIZONTAL standard (draft CEN proposal), www.ds.dk.
- [18] K.H. Esbensen, C. Paoletti, P. Minkkinen, Representative sampling of large kernel lots – I. Theory of Sampling and variographic analysis, TrAC, Trends Anal. Chem. 32 (2012) 154–165, <http://dx.doi.org/10.1016/j.trac.2011.09.008>.
- [19] K.H. Esbensen, C. Paoletti, P. Minkkinen, Representative sampling of large kernel lots – III. General considerations on sampling heterogeneous foods, TrAC, Trends Anal. Chem. 32 (2012) 179–184, <http://dx.doi.org/10.1016/j.trac.2011.12.002>.
- [20] P. Minkkinen, K.H. Esbensen, C. Paoletti, Representative sampling of large kernel lots – II. Application to soybean sampling for GMO control, TrAC, Trends Anal. Chem. 32 (2012) 66–178, <http://dx.doi.org/10.1016/j.trac.2011.12.001>.
- [21] P. Thy et al., On representative sampling and reliable chemical characterization in thermal biomass conversion studies, Biomass Bioenergy (2009), <http://dx.doi.org/10.1016/j.biombioe.2009.07.015>.
- [22] P.O. Minkkinen, K.H. Esbensen, Grab vs. composite sampling of particulate materials with significant spatial heterogeneity – a simulation study of “correct sampling errors”, Anal. Chim. Acta 653 (1) (2009) 59–70.
- [23] VIM, International Vocabulary of Metrology – Basic and General Concepts and Associated Terms, third ed., JCGM, 200, 2008.
- [24] VIM, 2010. Corrigendum (May 2010).
- [25] ISO 11074–2, 1998. Soil Quality – vocabulary. Part 2: terms and definitions related to sampling, International Organization for Standardization, Geneva.
- [26] EURACHEM, CITAC, 2000. Quantifying Uncertainty in Analytical Measurement. Eurachem, second ed. S.L.R. Ellison (LGC, UK), M. Rosslein (EMPA, Switzerland), A. Williams (UK).
- [27] EURACHEM, CITAC, 2012. Quantifying Uncertainty in Analytical Measurement. Eurachem, third ed. S.L.R. Ellison (LGC, UK), A. Williams (UK) (Eds.), ISBN 978-0-948926-30-3.
- [28] A.J.B. Cotta, J. Enzweiler, Classical and new procedures of whole rock dissolution for trace element determination by ICP/MS, Geostand. Geoanal. Res. 36 (2011) 27–50.
- [29] W. Zhang, Z. Hu, Y. Liu, L. Chen, H. Chen, M. Li, L. Zhao, S. Hu, S. Gao, Reassessment of HF/HNO₃ decomposition capability in the high-pressure digestion of felsic rocks for multi-element determination by ICP/MS, Geostand. Geoanal. Res. 36 (2012) 217–289.
- [30] P.J. Potts, Handbook of Rock Analysis. Viridian Publishing, 2003. ISBN0-9544891-3-6.
- [31] P. Minkkinen, Practical applications of sampling theory, Chemometr. Intell. Lab. Syst. 74 (1) (2004) 85–94.
- [32] K. Heydorn, K.H. Esbensen, Sampling and metrology, Accred. Qual. Assur. 9 (7) (2004) 391–396.
- [33] M. Ramsey, Contaminated land: cost-effective investigation within sampling constraints. In: P. Potts, M. West (Eds.), Portable X-ray Fluorescence Spectrometry: Capabilities for In Situ Analysis. RCS Publ., 2008, pp. 39–555. ISBN 978-0-85404-552-5.
- [34] M. Korpelainen, S.-P. Reinikainen, J. Laukkanen, P. Minkkinen, Estimation of uncertainty of concentration estimates obtained by image analysis, J. Chemometr. 16 (2002) 548–554.
- [35] P. Minkkinen, Evaluation of the fundamental sampling error in the sampling of particulate solids, Anal. Chim. Acta 1987 (196) (1987) 237–245.
- [36] V.R. Meyer, Measurement uncertainty, J. Chromatogr. A 1158 (2007) (2007) 15–24.
- [37] <http://www.impublications.com/tos-forum>, pp. 19–22.