Theory of sampling (TOS)—the missing link before analysis

Kim H. Esbensen[a] and Claas Wagner[b]
[a]Geological Survey of Denmark and Greenland (GEUS) and Aalborg University, Denmark. E-mail: ke@geus.dk
[b]Sampling Consultant. E-mail: cw@wagnerconsultants.com

WHAT is the meaning of analysing a non-representative sample? None—there is no meaning! Undoubtedly everybody agrees with the answer to the above provocative question. If a sample cannot be documented to be representative (of the lot/target material from where it was taken) it is a waste of time, effort and money to analyse it! This small argument pretty much sums up the background for the present new column in *Spectroscopy Europe*. While there is plenty of sampling going on in the world of science, technology and industry, much of it unfortunately cannot be characterised as anywhere near being representative, indeed it is rather oriented towards easy, not-too-expensive "sampling"—with the least effort. However, sampling can certainly be both practical and representative, all it needs is to follow the Theory of Sampling.

The ever increasing precision with which quantitative spectroscopy is able to analyse is often accompanied by an inversely decreasing analytical mass—the higher the analytical precision, the smaller the volume to be characterised. This is reflected by significant current efforts to increase the "effective sample presentation volume" (e.g. for simple quantitative NIR exemplified by the static petri dish being supplanted by a rotating dish, to be supplanted by the roll bottle, to be replaced by …). Though the degree to which this is manifested is significantly different for the various spectroscopic methods, the common issue is that the test portion continues to be but a very small fraction of the original target material (termed "lot" in the Theory of Sampling, TOS).

Precision cannot be traded for accuracy in the greater perspective, however. Irrespective of increased analytical precision (which is a matter solely related to the analytical method and equipment), the accuracy of an analytical result is related to the estimate of the analyte content of the complete lot from which the test portion has been derived: sampling accuracy is a matter solely related to the sampling process (and as will be shown, the analytical accuracy all but vanishes in this context). Barring trivial micro- and meso-scale laboratory examples from which no generalisation can be made, it is readily acknowledged that to come from a typical lot in science, technology or industry to the analytical aliquot, sampling must have been involved. It is not always equally readily comprehended, however, that the sampling rates involved are typically range orders of magnitude $1:10^6$, $1:10^9$ or $1:10^{12}$ (Figure 1). Sampling spanning such a significant number of orders of magnitude (mass : mass or volume : volume) to produce the analytical aliquot is far from a simple operation, because all materials are inherently ***heterogeneous*** (at some scale or other, which will be explained in detail in future columns). Sampling of heterogeneous materials is most emphatically not a simple materials handling issue! Heterogeneity is both the common denominator of all types of materials (of

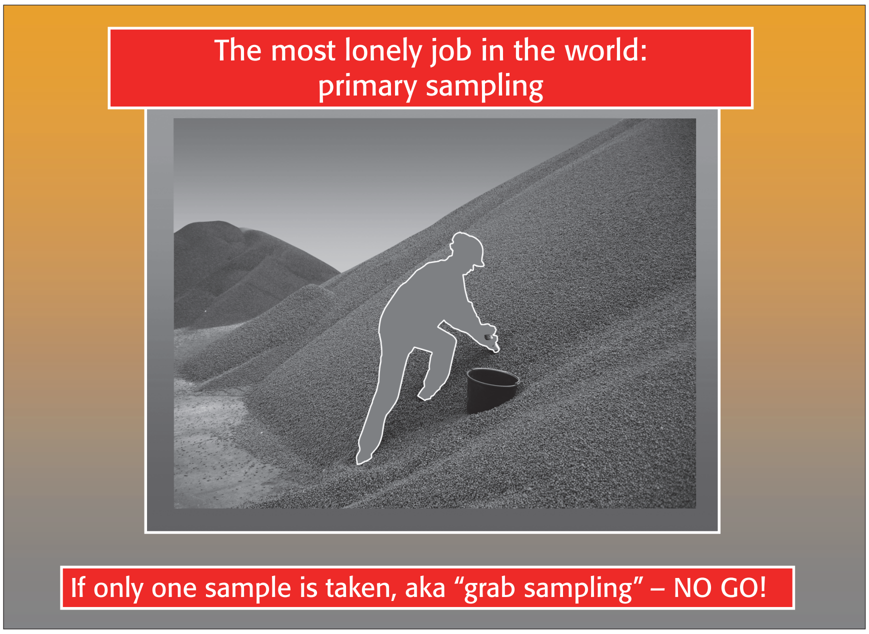

Figure 1. Where representativity starts, at the primary sampling stage—long before analysis. This generic illustration is a placeholder for the world's infinitely many different types of materials (lots). It also illustrates the crucial point that heterogeneity need not be visible…

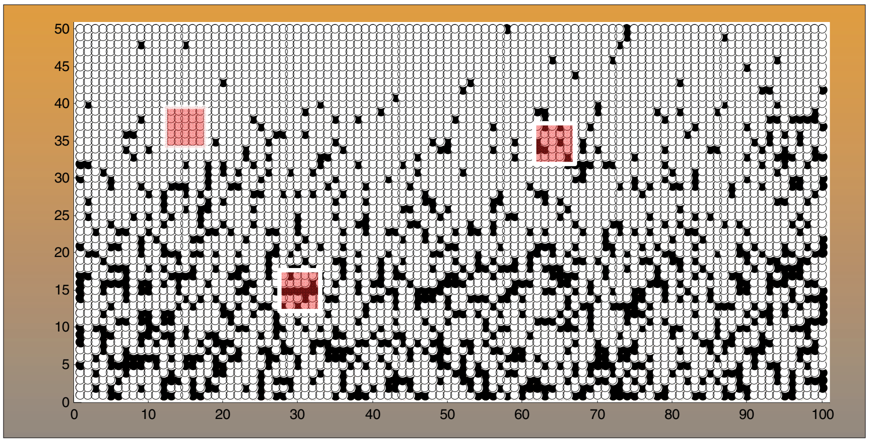

Figure 2. The arch-enemy of representative sampling: heterogeneity. A "grab sample" (a single event sampling) can never be representative of the whole lot.

all lots), as well as the arch-enemy of all sampling efforts, Figure 2.

There is not enough understanding and skill regarding the critical role of sampling. Here is the case in point. When confronting a lot (of any size, the issues below are scale-invariant) with a sampling purpose (with a sampling implement in hand, scoop, spear, whatever …), the following question appears fully justified: **"How big a sample is needed, in order for it to be representative?"** For lots of any size, the sampling issue, from almost all points of view, can easily overwhelm. Surprisingly, upon reflection guided by the Theory of Sampling (TOS), which is to be introduced in detail in this new column, this is, however, the *wrong question*, at the *wrong time*, in the *wrong place*!

The wrong question: because representativity is *not* related to sample size, but only to the sampling process—ascertaining whether a particular sample is representative or not can never be resolved by characterisation of the sample itself.

The wrong time: because this issue should have been thought through long before the actual sampling commences. This issue cannot be solved at the same time as when one is preoccupied with fixing the sample volume, i.e. sample size or mass.

The wrong place: because unless one has not already started learning a certain minimum of proper sampling principles, it is most likely that the focus is on taking just one sample as this is indeed the easiest—a procedure termed "grab sampling". However, the most important tenet of TOS is that grab sampling is *always wrong*—and that only composite sampling (multi-increment sampling) is able to guarantee representativeness. This column shall deal extensively with this type of sampling. In composite sampling, a sufficient set of individual increments covering the entire volume of the lot is essential; determining the numerical issue of "sufficient" is part of the definition of representativity, and is in fact the correct answer to the question of "how big a sample".

A roadmap for representative sampling

This column will deal with the principles of *representative sampling* (as opposed to mere mass reduction), presenting to the reader a comprehensive, necessary and sufficient platform with which to be able to perform, and document, representative sampling at all scales, for all types of materials (lots). Indeed the series of columns, when complete, will make up a complete professional competence basis for this task. The columns will try to treat all aspects of sampling, well-known or not and there will be many surprises as well. It will no doubt be advantageous to start with an initial roadmap to be filled in gradually. We outline below such a roadmap (given the space available), in the briefest possible overview format.

Section 1 (introduction to TOS)

- Why sampling? (why materials handling is not sampling)
- The Theory of Sampling (TOS)—fundamental principles and definitions
- Sampling terminology (the tower of Babel)
- The sampling target, the lot (lot dimensionality: 0, 1, 2, 3 dimensional lots)
- Heterogeneity—the arch-enemy
- Representativity—a formal definition
- Representativity is solely a characteristic of the sampling process
- Representative sampling is always a multi-stage process
- Sampling errors
- Correct and incorrect sampling errors
- Process sampling errors

Section 2 (TOS applications)

- Sampling Unit Operations (SUOs)
- Special focus on mass reduction (one of the SUOs)
- Heterogeneity characterisation
- The Replication Experiment (RE) for stationary lots
- Variographic analysis (a first preview)
- Process sampling (sampling of moving and stationary 1-D lots)
- Variograms in detail
- Sampling in the 2-D plane (what is so special about 2-D?)
- Four Quality Criteria to ensure representative sampling
- TOS vs Measurement Uncertainty (MU)—a call for integration
- The analytical bias is constant—but the sampling bias is not

Section 3 (History, further information, guidelines, standards, sampling community)

- International standards, guidelines, norm-giving documents … …
- *DS 3077 Horizontal*—A unified international standard for representative sampling
- Pierre Gy—a monumental scientific *oevre*
- World Conference on Sampling and Blending (WCSB)
- Sampling Hall of Fame/Shame (many instructive case histories from which to learn)

- TOS literature
- *TOS forum*

Section 4 (Readership interaction)

- Practical sampling examples supplied by the readership
- Case histories supplied by the readership (YOU!)

What comes before analysis?

The following scene-setting is valid for all types of subsequent analysis, spectroscopic or otherwise.

In one sense it is *all* about what comes before analysis, in the sense of the provocative question stated in the introduction: what is the meaning of analysing any sample if it cannot be documented to be representative? How does one achieve and document that a specific sample is indeed representative of the target material?

Here comes the first foray of surprises in this new column. It is *not* possible to document that a *specific* sample is representative by any known method, approach or activity—be this analytical, data analytical, statistical or otherwise. It is not possible to discern the status of any sample from any type of inspection of the sample itself. All specific samples, when observed in isolation, only allow one characterisation—they constitute a very small, mass-reduced (representative or not) fraction of the lot. But mass reduction alone has nothing to do with representativity. It is only the specific *sampling process* with which a sample was extracted that can be designated as representative, or not, according to certain criteria which will be presented later. Thus, in general there must be sampling from a lot; there must in all likelihood also be sub-sampling (additional mass-reduction steps) in order to furnish the ultimate goal of sampling, the analytical aliquot. It would be fatal to ignore these pre-analysis steps and only focus on the final analytical activity. As shall become clear, the potential sum total of the sampling and sub-sampling errors typically dominate in the total uncertainty budget compared to analysis alone. It is not unusual to surpass the analytical uncertainty with factors 10–20–50 … depending on the degree of heterogeneity of the material sampled and on the degree to which the sampling process has been purged of the complement of sampling errors. These special Sampling Columns will also address how to characterise and quantify heterogeneity and even more important how to counteract heterogeneity by applying TOS principles of sampling correctness.

Designing a sampling process is a futile undertaking if not related to the material heterogeneity encountered. The Theory of Sampling (TOS) includes specific guidelines for how to estimate lot heterogeneity, both for stationary as well as for dynamic (moving) lots. At the same time this is accomplished, one will actually be able to optimise the specific sampling process involved.

There exists a certain minimum competence basis regarding the Theory of Sampling (TOS) that must be acquired by any professional sampling operator or personnel dealing with samples or analytical aliquots, otherwise the prospects of documentable representativeness will be forfeited. There is plenty of literature, at all conceivable levels, that will allow anybody interested to acquire this competence with no exceptions, our first offer being References 1–6.

The next Sampling Column will introduce the fundamental definitions and principles of the Theory of Sampling, especially focusing on lot dimensionality, a TOS term defining the geometrical dimensions as well as the effective number of dimensions involved in sampling.

Welcome!

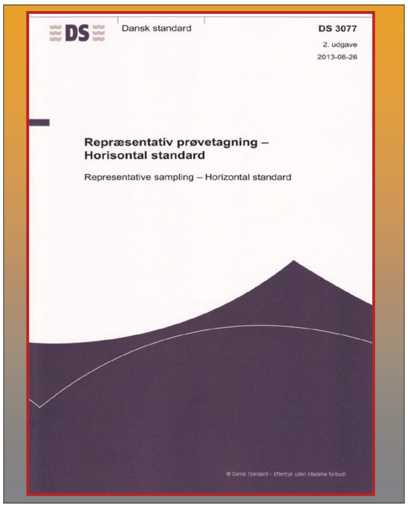

Figure 3. The world's first matrix-independent ("horizontal") standard on representative sampling[2] provides the most concise, yet comprehensive introduction to the TOS.

References

1. www.impublications.com/tos-forum
2. K.H. Esbensen, DS 3077. *Representative sampling—Horizontal Standard*. Danish Standards (2013). www.ds.dk
3. K.H. Esbensen and L.P. Julius, "Representative sampling, data quality, validation—a necessary trinity in chemometrics", in *Comprehensive Chemometrics*, Ed by S. Brown, R. Taule and B. Walczak. Elsevier, Vol. 4, pp. 1–20 (2009). doi: http://dx.doi.org/10.1016/B978-044452701-1.00088-0
4. K.H. Esbensen and C. Wagner, "Theory of Sampling (TOS) vs. Measurement Uncertainty (MU)—a call for integration", *Trends Anal. Chem.* **57,** 93–106 (2014). doi: http://dx.doi.org/10.1016/j.trac.2014.02.007
5. K.H. Esbensen and C. Wagner, "Why we need the Theory of Sampling", *The Analytical Scientist* #21, 30–38 (October 2014).
6. C. Wagner and K.H. Esbensen, "A critical review of sampling standards for solid biofuels—Missing contributions from the Theory of Sampling (TOS)", *Renew. Sust. Energ. Rev.* **16,** 504–517 (2011). doi: http://dx.doi.org/10.1016/j.rser.2011.08.016

Kim H. Esbensen originally trained as a geologist/geochemist, but it was 30 years before he actually worked in a geoscience institution (The Geological Survey of Denmark and Greenland). In between he established two research groups dealing with PAT and chemometrics. He found a third love, scientifically speaking, some 15 years ago, when he met the Theory of Sampling (TOS), and the field of representative sampling has occupied his career ever since. Kim is specifically interested in the interaction between process and material heterogeneity, representative sampling and augmented measurement uncertainty.

Originally trained as an economist, **Claas Wagner** realised that his real interests were with environmental and energy related topics and therefore continued his education in this direction. Sustainable resource management, emission reduction procedures and energy efficiency issues have all one common ground: decisions need to be based on valid data. This led to Claas' PhD on representative sampling and data analysis for quality monitoring in large-scale combustion plants. Currently Claas combines his fields of interest, working as a consultant for various industries providing quality assurance approaches. Throughout all of this reigns representative sampling.