LECTURE

The critical role of representative sampling before analysis, NIR or otherwise

Kim H. Esbensen

ke@geus.dk

KHE Consulting, CAMO Software, Aalborg University, Denmark, Geological Survey of Denmark & Greenland, Telemark University College, Norway

Analytical results are meant to be representative of the original target (lot) from which they are derived. The specific analytical performance notwithstanding, the sample size contrast between lot (typically *kg-ton*) and analytical aliquot (*mg-g*) is of the order of $1/10^6$ to $1/10^9$ *or higher*. The critical success factor for analysis, NIR or otherwise, is therefore that sampling preceding analysis must be fully representative (and fully documentable) at all stages along the full "lot-to-analysis" process. The principles governing this demand are all are covered by the Theory of Sampling (TOS). The analytical, chemometric, PAT and MSPC realms have been presented with reasonable efforts of introducing TOS in the last 5+ years, and has also seen a new international standard for sampling, DS 3077 (2013). Little more needs therefore to be added here regarding sampling of *stationary lots* (0-, 2- and 3-D lots). All primary lots and materials to be samples are *heterogeneous*, at all scales. The key feature governing sampling effectiveness is to what degree the sampling process is able to *counteract* the effects of heterogeneity, lest a fatal *sampling bias* will lead to unnecessarily inflated sampling errors, which cannot be corrected for - the sampling bias is *inconstant*. Whereas an analytical bias can always be compensated for, because it can be assumed to be constant, the sampling bias cannot under any circumstances be subject to a similar correction; this is perhaps the most striking new insight provided by TOS to the analytical sciences in general, to NIR in particular.

This keynote lecture mainly focuses on the specific principles governing representative TOS *process sampling* (1-D lots, dynamic lots), whether by physical sample extraction or by so-called "sampling-free" approaches, i.e. by signal acquisition from PAT sensors (NIR, other modalities). It is here emphasized that there is a complete duality between these two cases and that there is nothing special and certainly nothing superior regarding PAT (nor handheld instrumentation). In fact "sensor sampling" is subject to the exact same error potential and demands as traditional sample extraction, both in the process realm and everywhere else, and has indeed seen a marked proportion of insufficiently justified assumptions and implementations

(examples will be presented). But this duality also makes it possible to offer a coherent, unifying approach to representative sampling of all type of lots, whether stationary and dynamic, TOS.

This lecture presents a new approach from 1-D TOS, *variographics*, which is able to decompose the total observed process variation into: i) total measurement system error and ii) true (unmasked) process variability. The former has in very many PAT and MSPC applications been a source of significant degradation of true process observability. We here present this approach for on-line total measurement system quality control & assurance, VARIOQC, with examples histories the pharmaceutical and petrochemical process industry sectors.