Analytica Chimica Acta 653 (2009) 59–70

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Grab vs. composite sampling of particulate materials with significant spatial heterogeneity—A simulation study of "correct sampling errors"

Pentti O. Minkkinen*, Kim H. Esbensen

ACABS Research Group, Aalborg University, Campus Esbjerg, Niels Bohrs Vej 8, DK-6700, Denmark (AAUE), Finland

ARTICLE INFO

Article history:
Received 6 April 2009
Received in revised form 31 August 2009
Accepted 31 August 2009
Available online 2 September 2009

Keywords:
Theory of Sampling (TOS)
Process sampling
Material heterogeneity
Sampling process
Simulation
Sampling uncertainty

ABSTRACT

Sampling errors can be divided into two classes, incorrect sampling and correct sampling errors. Incorrect sampling errors arise from incorrectly designed sampling equipment or procedures. Correct sampling errors are due to the heterogeneity of the material in sampling targets. Excluding the incorrect sampling errors, which can all be eliminated in practice although informed and diligent work is often needed, five factors dominate sampling variance: two factors related to material heterogeneity (analyte concentration; distributional heterogeneity) and three factors related to the sampling process itself (sample type, sample size, sampling modus). Due to highly significant interactions, a comprehensive appreciation of their combined effects is far from trivial and has in fact never been illustrated in detail. Heterogeneous materials can be well characterized by the two first factors, while all essential sampling process characteristics can be summarized by combinations of the latter three. We here present simulations based on an experimental design that varies all five factors. Within the framework of the Theory of Sampling, the empirical Total Sampling Error is a function of the fundamental sampling error and the grouping and segregation error interacting with a specific sampling process. We here illustrate absolute and relative sampling variance levels resulting from a wide array of simulated repeated samplings and express the effects by pertinent lot mean estimates and associated Root Mean Squared Errors/sampling variances, covering specific combinations of materials' heterogeneity and typical sampling procedures as used in current science, technology and industry. Factors, levels and interactions are varied within limits selected to match realistic materials and sampling situations that mimic, *e.g.*, sampling for genetically modified organisms; sampling of geological drill cores; sampling during off-loading 3-dimensional lots (shiploads, railroad cars, truckloads etc.) and scenarios representing a range of industrial manufacturing and production processes. A new simulation facility "SIMSAMP" is presented with selected results designed to show also the wider applicability potential. This contribution furthers a general exposé of all essential effects in the regimen covered by "correct sampling errors", valid for all types of materials in which non-bias sampling can be achieved.

1. Introduction

Any reliable approach to practical, representative sampling must be based on the Theory of Sampling (TOS), as concerns both stationary lots as well as process sampling. In recent years there have been a number of presentations of all essential aspects of TOS within chemometrics and analytical chemistry, eliminating the need for yet another brief of the Theory of Sampling (TOS) here; comprehensive references are widely available on theory [1–4] and applications [5–15]. Several international guides are also available, *e.g.*, Refs. [16–18].

Sampling errors can be divided into two classes called incorrect sampling (ISE) and correct sampling errors (CSE). Incorrect sampling errors arise from incorrectly designed sampling equipment or procedures. Correct sampling errors are due to the heterogeneity of the material in sampling targets. Within this framework elimination of the incorrect sampling errors always forms the prime objective, lest sampling is marred by a fatal sampling bias, ibid.

In the present contribution we shall assume that all ISE have been fully eliminated, leaving only the correct sampling errors for detailed study. In this regimen, proper understanding of the most dominating factors, the fundamental sampling error (FSE) and grouping and segregation error (GSE) is essential. Sampling is then governed by the following facts:

1. All materials are heterogeneous, at all scales.
2. Any sampling process produces its own sampling errors.

* Corresponding author at: Lappeenranta University of Technology, P.O. Box 20, FIN-53851, Lappeenranta, Finland. Tel.: +358 40 504 9413.
E-mail addresses: Pentti.Minkkinen@lut.fi, Pentti.Minkkinen@pp.inet.fi (P.O. Minkkinen).

0003-2670/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.aca.2009.08.039

3. Many types of technological equipment currently in use are flawed in design w.r.t. TOS's criteria for representative sampling thus generating sampling bias.

All incorrect sampling errors (ICS) can be eliminated, in principle, although in practice informed and diligent work is often needed [1–9], after which there are five dominating factors influencing sampling uncertainty, *viz.*, two related to material heterogeneity: analyte concentration; type of heterogeneity and three related to the sampling process itself: sample size, sampling modus (random, stratified, systematic sampling) and sample types (single increment, or grab vs. several increments combined into a composite sample). All five factors will be included in the simulations presented below.

While the sampling literature is ripe with many partial illustrations of one or more subsets of these factors influencing the practical sampling uncertainty (sampling variance) [1–19] their combined effects resulting from significant interactions have not at all been explored sufficiently. It has long been a desire to be able to elucidate this theme for all professions in need of representative sampling of stationary as well as moving lots. A didactic illustration has now been accomplished as the first results from a comprehensive simulation study based on all five major influencing factors, which have been explored according to the following specifications:

F1. Analyte concentration, 5 levels [1%, 5%, 10%, 20%, 40%]

F2. Heterogeneity type, 7 levels [randomly distributed analyte particles; analyte particles in randomly distributed clusters of 2, 4, 6, 16, and 32 particles; segregated, grading lot]

F3. Sample size, 6 levels [consisting of 1, 5, 10, 25, 50, 75 and 100 consecutive particles]

F4. Sampling type, two levels [composite sampling *vs.* single increment, or grab sampling] with the same total sample size

F5. Sampling mode, three levels [random-, stratified- and systematic sampling]

ISE[1] are suppressed completely in the present experiments, in order to focus on the specific materials-related factors and the intrinsic sampling process factors as a function of the above mentioned five experimental factors. ISE result from incorrectly designed sampling equipment or from failure to obey the rules of bias-free sampling [1–8]. In this context, the present contribution offers a singularly detailed look into the effects covered by the two so-called "correct sampling errors", FSE and GSE, and their interaction with a relevant range of typical sampling process types.

Factors 4 and 5 are qualitative factors. Factors 1–3 are continuous. Levels were selected to cover typical ranges met with in practical sampling; also they span the experimental space maximally. The total number of all possible interactions between the above five factors (with between 2 and 7 levels) will lead to an unmanageable plethora of results, far too much for individual presentation. We shall here restrict illustrations to a number of well-spanning, typical cases, as well as presenting relevant summaries of many more cases than what can be individually visualized.

1.1. Sampling terminology in brief

Lot – is the sampling target from which increments (see below) are cut in order to estimate its properties, *e.g.*, the lot mean value, a_L, or the sampling variance of the mean, $s^2_{a_L}$.

[1] ISE consists of IDE, increment delimitation error, IME: increment materialization error and IPE: increment preparation error [1–4].

Sample – A "representative" sample is the result from a "correct" sampling process ("correct" is explained below). An incorrectly sampled portion of the lot is termed a specimen.

Specimen – A specimen is made up of material extracted from the lot in any incorrect fashion. TOS puts critical emphasis on the difference between a sample and a specimen. From a specimen one cannot draw valid conclusions concerning the properties of the whole lot.

Correctness – "Correctness" is a qualifying term designating that (all) incorrect sampling errors (ISE, see text above) have been eliminated in the sampling process.

Increment – An increment is the portion of the material extracted from the lot as the first step of the sampling process. An increment can be a sample by itself or used for compositing.

Composite sample – A composite sample is by definition made up of several, independently taken increments.

Representative sample – sample extracted using only 'correct' procedures, with the connotation that the sample has an uncertainty that is *fit for intended use*. Incorrectly extracted samples cannot be representative.

Sampling types

- *Grab sampling* – Any "lump" of material from the lot, "taken in one single operation". Grab sampling in this paper means that all primary increments are processed and analyzed separately.
- *Composite sampling* – a number (N) of primary increments are combined into one aggregate (or bulk) sample before further processing and analysis.

Sampling modes (or strategies)

- *Random sampling* – Sampling locations are randomly assigned within the lot, either along the time axis or geometrical axes, depending on case
- *Stratified random sampling* – The lot is divided into sub-lots (called *strata*) before sampling. Within each stratum, increments, or samples (grab or composite) are selected randomly.
- *Systematic sampling* – Samples are cut at equal distances (lags) along the time (or geometric) axes of the lot.

Cluster – A special form of segregation, where particles of interest (analyte particles) form localized groups within the matrix of the lot (short range variability).

Fragment – In TOS a "fragment" is the smallest physically separable unit that is not physically affected by the sampling process. Usually what is thought of is a granule, a grain (or even a molecule), but TOS' definition includes both natural particles as well as true fragments resulting from crushing etc. as function of the sampling process. This definition allows one to consider both unaffected as well as fragmented particles by the comprehensive cover term "fragments". In this way it is possible to treat all types of sampling process in a unified manner.

Group – In TOS a group of fragments consists of local, spatially correlated fragments which act as a coherent unit during the sampling procedure. In practical sampling, a group either forms an increment, a specimen or a sample. Note that the particular group-of-fragments, which occupies the sampling tool volume, constitutes the most interesting element in sampling, the increment.

2. Scope and disposition of simulations

All simulations are carried out in MATLAB 7.3 using SIMSAMP (Simulated Sampling), a new general purpose sampling simulation facility designed for illustrating the effects of interaction between