Sampling of particulate materials with significant spatial heterogeneity - Theoretical modification of grouping and segregation factors involved with correct sampling errors

Fundamental Sampling Error and Grouping and Segregation Error

Minkkinen, Pentti O.; Esbensen, Kim H.

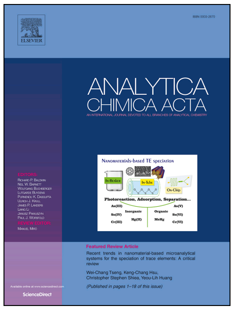
Published in:
Analytica Chimica Acta

DOI (link to publication from Publisher):
10.1016/j.aca.2018.10.056

Publication date:
2019

Document Version
Accepted author manuscript, peer reviewed version

Citation for published version (APA):
Minkkinen, P. O., & Esbensen, K. H. (2019). Sampling of particulate materials with significant spatial heterogeneity - Theoretical modification of grouping and segregation factors involved with correct sampling errors: Fundamental Sampling Error and Grouping and Segregation Error. *Analytica Chimica Acta*, *1049*, 47-64. https://doi.org/10.1016/j.aca.2018.10.056



Sampling of particulate materials with significant spatial heterogeneity - Theoretical modification of grouping and segregation factors involved with correct sampling errors: Fundamental Sampling Error and Grouping and Segregation Error

Pentti O. Minkkinen, Kim H. Esbensen

PII: S0003-2670(18)31298-4

DOI: https://doi.org/10.1016/j.aca.2018.10.056

Reference: ACA 236362

To appear in: *Analytica Chimica Acta*

Received Date: 8 June 2018

Revised Date: 5 October 2018

Accepted Date: 24 October 2018

Please cite this article as: P.O. Minkkinen, K.H. Esbensen, Sampling of particulate materials with significant spatial heterogeneity - Theoretical modification of grouping and segregation factors involved with correct sampling errors: Fundamental Sampling Error and Grouping and Segregation Error, *Analytica Chimica Acta*, https://doi.org/10.1016/j.aca.2018.10.056.

Reducing contributions to total sampling variance from the Grouping and Segregation Error (*GSE*)

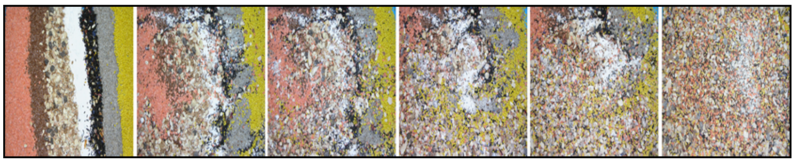
$$s^2(GSE) = z \cdot y \cdot s^2(FSE), y \approx \frac{1}{n_{inc}} = \frac{m_{inc}}{m_{samp}}$$

Segregation factor, z

Grouping factor, y (unaffected by mixing - reduced only by selecting smaller increments)

$z > 1$

$z \approx 1$



m_{samp} = mass of analyzed samples, m_{inc} = mass of increment

n_{inc}= number of increments in analyzed samples

ACA-18-1638Rev.Highlighted

Sampling of particulate materials with significant spatial heterogeneity - Theoretical modification of grouping and segregation factors involved with correct sampling errors: Fundamental Sampling Error and Grouping and Segregation Error

Pentti O. Minkkinen*[a], Kim H. Esbensen[b]

*Corresponding author at: Lappeenranta University of Technology, P.O. Box 20, FI-53851 Lappeenranta, Finland. Tel. +358 40 504 9413

Professor emeritus

E-mail: Pentti.Minkkinen@lut.fi

[a]Sirpeka Oy, Lampisuonkatu 9, FI-53850 Lappeenranta, Finland,

E-mail: pentti.minkkinen@sirpeka.fi

[b]KHE Consulting, Aldersrogade 8, DK-2100 Copenhagen, Denmark

E-mail: khe.consult@gmail.com

Adjunct professor, Aalborg University (AAU), Denmark
Adjunct professor, Geological Survey of Denmark and Greenland (GEUS)
Professeur associé, Université du Québec à Chicoutimi (UQAC), Quebec (2018-2020)
Guest professor University of Southeast Norway (USN)
Guest professor (2018) Recinto Universitario de Mayaguez, Puerto Rico

Declarations of interest: none

Sampling of particulate materials with significant spatial heterogeneity - Theoretical modification of grouping and segregation factors involved with correct sampling errors: Fundamental Sampling Error and Grouping and Segregation Error

Pentti O. Minkkinen, Kim H. Esbensen

Abstract

There has been an extensive abuse of Gy's Formula during the entire history of applied TOS (Theory of Sampling), it being applied too liberally to almost any aggregate material conceivable for many material classes of extremely different compositions with significant (to large, or extreme) fragment size distribution heterogeneity, for example many types of municipal and industrial waste materials. This abuse regimen is for the most part characterized by lack of fundamental TOS competence and the historical context of Gy's formula. The present paper addresses important theoretical details of TOS, which become important as sampling rates increase at the conclusion of the full 'lot-to-analysis sampling pathway regarding finer details behind TOS' central equations linking sampling conditions to material heterogeneity characteristics allowing the estimation of Total Sampling Error (*TSE*) manifestations. We derive a new, complementary understanding of the two conceptual factors, *y* the *grouping factor* and, z, the *segregation factor*, intended to represent the *local* (increment scale) and *long-range* (increment to lot-scale) heterogeneity aspects of lot materials, respectively. We contrast the standard TOS exposé with the new formulation. While the phenomenological meaning and content of the new proposed factors (*y* and *z*) remains the same, their numerical values and bracketing limits are different with *z* now representing more realistic effects of liberation and segregation combined. This new formulation makes it easier to get a first comprehensive grasp of TOS' dealings with sampling of significantly heterogeneous materials. We believe this may present a slightly easier path into the core issues in TOS when sampling and sub-sampling gets closer to the final aliquot scale.

KEY WORDS: *Sampling theory; Sampling uncertainty; Heterogeneity; Segregation; Sampling Errors*

1 Introduction

1.1 Summary of earlier studies

Theory of Sampling (TOS) distinguishes two classes of sampling errors, termed incorrect and correct sampling errors respectively. Incorrect sampling errors (*ISE*) arise from wrongly designed sampling equipment or inferior sampling procedures which all can be significantly reduced and/or eliminated in practice, however, although informed and diligent work is often needed. The remaining correct sampling errors (*FSE*, *GSE*) arise from the interaction between a particular sampling procedure and specific the heterogeneity of the target material when the sampling process is correctly designed and operated following all of TOS' pertinent rules. In an extensive study Minkkinen & Esbensen [1] investigated the influence of five factors that dominate the sampling variance in this case: two factors relate to material heterogeneity (analyte concentration and compositional heterogeneity/distributional heterogeneity) and three factors that relate to the sampling process itself (sample type, sample size, sampling modus). Significantly, heterogeneous materials are well represented by the two first factors, while all practical sampling process characteristics can be understood as combinations of the latter three. Extensive *in silico* simulations were presented based on an experimental design that varied all five factors systematically. A wide array of repeated simulated sampling campaigns was run and the results were expressed as illustrations showing the pertinent effects as lot mean estimates and the associated Root Mean Squared Errors (*RMSE*), covering a range of typical combinations of materials' heterogeneity and often used sampling procedures as applied in science, technology and industry. Factors, levels and interactions were varied within limits selected to match *realistic materials* and sampling situations that represent important cases, 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 simulation facility "SIMSAMP" (MATLAB) was presented, with comprehensive results designed to show also a wider applicability potential. The study concentrated on estimating the effects of these heterogeneity types on the "correct sampling errors", aimed at being valid for all types of materials in which non-bias sampling can be achieved.

1.2 Scope of present paper

The present paper addresses some important issues of heterogeneity, notably a re-evaluation of the theoretical derivations behind TOS' equations linking sampling conditions to the inherent material heterogeneity characteristics, allowing the estimation of Total Sampling Error (*TSE*) manifestations. In this context we derive a new, complementary understanding of the two conceptual factors in TOS, Y the *grouping factor* and, Z, the *segregation factor*, (y and z with new definition) intended to represent the two most important processes responsible for the *local*, and the *long-range* heterogeneity aspects of lot materials. In this development, new insights into certain finer details of Pierre Gy's [2, 3, 4] original theoretical derivations come to light. We contrast the standard TOS exposé with the new formulation. The boundary conditions for the present foray is that the phenomenological meaning and content of both the Y and Z factors remains the same, but their numerical values and bracketing limits are now different. We submit that the new formulation makes it easier to get a first comprehensive grasp of TOS' dealings with sampling of significantly heterogeneous materials, especially as sampling rates increase at the conclusion of the full 'lot-to-analysis sampling pathway. The new understandings presented here apply when sample masses approach ~2-10% of the penultimate sub-sample masses in the terminal stages of the full lot-to-aliquot pathway, e.g. as when a spatula is used for delivering the final analytical aliquot mass, or similar in the process sampling and PAT domains.

2 TOS – a brief

2.1 Homogeneity – heterogeneity – sampling errors

Perfect homogeneity, i.e., a spatially 'randomly distributed lot', is an ideal non-existing property. Heterogeneity is the rule - all naturally occurring and manufactured materials are heterogeneous; it is only a matter of degree. Heterogeneity is manifested at all scales of interest in a lot; as grouping and segregation phenomena.

Material heterogeneity is the source of the correct sampling errors (CSE) and will influence the quantitative expression of the incorrect sampling errors (ISE) (preparation error being the only exception, see further below). Gy [3, 4] gave heterogeneity a comprehensive, detailed mathematical expression for the first time, although many incipient partial attempts are on record, see Minnitt: "Pierre Gy Oration" at WCSB8, Perth, 2017 [5]. A superior source from which to start learning about these complex historical matters is Pierre Gy's own review of the development history of TOS [6 – 9] and recent excellent reviews of the intricacies of Gy's mathematical developments Pitard and François-Bongarçon [10] and Matheron [11]. These are highly recommended for the reader interested in the full theoretical depth and power of TOS.

Here we proceed largely without this deep mathematical-statistical background. However, it was necessary to present TOS' unifying and sufficient minimum of Governing Principles (GP) and Sampling Unit Operations (SUO), which are sufficient for a practical understanding allowing the present objective to be fully understood (see Table 1). In order to prepare the way for the developments below, this recent compact summation of TOS as a set of axiomatic principles and unit operations is, therefore, presented by fiat; but see DS3077 [12] and Esbensen & Wagner [13] for details.

For the treatment below we only need to remind of another of Gy's inspirations, that of calling all original lot units (think of grains, particles etc.) as well as the resulting cascade of fragments hereof produced by accidental fragmentation as part of the impact from the sample preparation process itself, as *fragments*. Gy thus termed both the original units and their fragmented offspring *fragments,* thereby being able to treat the complete lot heterogeneity realm in a unifying manner: sampling (usually in the form of incremental extraction) can on this basis be dealt with complete unity irrespective of the specific nature and status of the complement of *fragments* making up the lot.

TABLE 1.

2.2 Theoretical nexus of TOS

In deriving the complete theoretical foundation for TOS, which took Pierre Gy 25 years, everything starts with developing a comprehensive understanding, in full mathematical statistical dressing, of the phenomenon of *heterogeneity*. Gy strived quite a bit with several different approaches, at first trying to formulate principles for describing heterogeneity as a function of two fundamental features ('factors'), *wiz,* the distribution of compositional differences between lot ‘units’, and the distribution of the inherent unit (fragment) size distribution. However, Gy was never fully satisfied with this approach; see Gy [6-9]. Matters came to a head in 1974, at a course given in Brazil, when a suggestion was offered to try to treat these independent distributions as just one, expressed as the product of the two factors involved (a typical ‘engineering solution’, somewhat outside the deeply theoretical quest Gy was on at the time). However, this turned out to be the pivotal incentive Gy needed, a breakthrough, and he coined the concept of the *heterogeneity contribution*, *h*, which ties together the composition and mass aspects of heterogeneity in a comprehensive tractable fashion.

The heterogeneity contribution is a quantity that can be assigned to - in fact it *represents* the heterogeneity arising from, both single fragments, h_i and increments (groups-of-fragments), h_n.

The general definition is as follows (for fragments, index *i*; for groups, index *n*):

$$h_i = \frac{a_i - a_L}{a_L} \cdot \frac{m_i}{\bar{m}} \tag{1}$$

$$h_n = \frac{a_n - a_L}{a_L} \cdot \frac{m_i}{\bar{m}_n} \tag{2}$$

In which $\bar{m}$ and $\bar{m}_n$ are the average mass of fragments, or average mass increments consisting of groups fragments, as the case may be and a_L is the weighted average content of the lot weighted with the pertinent fragment /group masses. Each heterogeneity contribution, h_i, *carries* a contributing fraction of the total heterogeneity of the lot.

By formulating all further theoretical developments in TOS on this basis, a simplification was obtained that allowed Gy to finish the comprehensive theoretical work in essentially just one more year. Gy himself spoke about the fulfillment of the Theory of Sampling in the year 1975. This year thus marks the divides into TOS’ gestation period (1950-1975) and the subsequent period of equal length 1975-2000, broadly speaking, the latter period mainly devoted to disseminating TOS to as many scientific and technological fields and industrial sectors as possible.

There exist a bewildering number of different classes and types of materials with their individual and common aspects of *heterogeneity*. Before Gy's theoretical breakthrough, it made very little intuitive sense that all the world's myriads of extremely different materials could ever be described by just one conceptual and mathematical framework. Yet this was the reality before the advent of the heterogeneity contribution. The reader is referred to the autobiographical account of the scientific development history of TOS in the words of its originator himself Gy [6-9].

In all material lots, individual *fragments* import their unique share to the total lot heterogeneity; likewise if the point-of-view is at the scale of *groups-of-fragments* (*groups* for short); groups are identical with *increments*, the practical sampling units.

The power of TOS is related to three scales only, fragment scale, increment scale and the scale (size) of the whole lot, which is viewed as made up by the totality of the heterogeneity contributions from either of these smaller-scaled units[1]. Thus, in order to be able to describe the heterogeneity of all types of lots (indeed all sizes of lots as well), one only needs these fundamental scales where the heterogeneity contribution concept plays out its role.

---

[1] N.B. A sample can either be a single increment or a composite sample made up of several increments covering the whole or parts of the lot volume. When no confusion can arise, below the term 'sample' is used without loss of generality.

At the end of this theoretical development, Gy was able to encapsulate the central aspects of the key relationships between practical sampling conditions and the material heterogeneity in just three master equations presented below:

1. "Gy's formula" for the variance of the Fundamental Sampling Error, originally published 1955 [2].
2. The central relationship between distribution (DH_L) and constitution (CH_L) heterogeneity, which can be expressed as: $DH_L = var(h_n) = Y \times Z \times CH_L = Y \times Z \times var(h_i)$
3. Another, slightly more complicated relationship: $DH_L = \frac{1+Y\,Z}{1+Y}\,CH_L$

It is possible to derive three of the four practical Sampling Unit Operations (SUO) from just two of these equations and indeed several of the Governing Principles as well, see e.g. Esbensen & Wagner [13], Esbensen [14].

2.3 TOS' definition of CH_L and DH_L

Gy's insight regarding how to describe all aspects of heterogeneity necessary for dealing with sampling (via the heterogeneity contribution concept), is that only the fragment scale and the group scale are needed. He realized that there are only two *kinds* of heterogeneity manifestations needed for a full theoretical treatment, wiz. Composition Heterogeneity, CH_L, and Distribution Heterogeneity, DH_L. Conceptual simplification resulted from the realization that CH_L and DH_L account for two *complementary* aspects of the same lot total heterogeneity, but that these play out their role at different scales, wiz. the fragment scale (CH_L) and the group scale (DH_L) respectively. CH_L and DH_L are conceptual *complements* accounting for the total lot heterogeneity but *as seen* from these two different observation scales. Nevertheless, these two components cannot be physically identified and separated from one-another. These heterogeneity components are both needed as theoretical components that play their role in TOS' treatment of heterogeneity and its influence on practical, empirical sampling.

CH_L and DH_L makes it possible to use the underlying mathematical formulation to *derive* the four SUO's with which all of practical sampling is carried out, always governed by the six GPs. This achievement forms the basis for a unified approach of practical sampling of the bewildering number of different materials and lots met with in science, technology and industry. This systematic approach to sampling has always been implicit in the original mathematical framework, but never formulated and systematized fully as laid out in Table 1.

In keeping with this two-scale approach, Gy defined the sum-total heterogeneity impact from all fragments, <u>and</u> from all virtual groups in the lot, (see below), by the *same* statistical formalism.

CH_L: The constitution heterogeneity of *all* N_F lot fragments is defined:

$$CH_L = \sigma^2(h_i) = (\sum_i h_i^2)/N_F \qquad (3)$$

The essential feature is that CH_L is defined as the variance of <u>all</u> heterogeneity contributions from <u>all</u> fragments that together make up the <u>whole lot</u>.

Regarding DH_L, the scale of observation is changed from that of fragments to that of groups-of-fragments (groups) but otherwise the argument is identical.

DH_L: The distribution heterogeneity, DH_L, is defined as the variance of the heterogeneity contributions between all increments (groups), $s^2(h_n)$; there are N_G potential groups making up the whole lot.

$$DH_L = \sigma^2(inc) = \sum_n h_n^2/N_G \qquad (4)$$

Again, the whole can also be viewed as being made up of <u>all</u> groups.

TOS now further derives a key interrelationship between CH_L and DH_L, by invoking two phenomenological factors, Y and Z, the grouping factor and the segregation factor.

2.3.1 Y **and** *Z*

The definition and meaning of Y and Z in TOS is as follows:

Y is the *grouping factor*, a conceptual factor used in the theoretical development to quantify the degree of *local* heterogeneity, i.e., the effect of *grouping*. However, it is also physically identifiable as a measure of the *size* of the sampling unit, the increment (group of fragments) but expressed in a somewhat surprising unit (not in a mass unit), see further below.

After a series of attempts at reducing the heterogeneity characterizing equations with various 'simplifying expressions', Gy succeeded with the following definition of Y (the formalism of this ratio originates as the so-called 'finite lot' correction found in statistics):

$$Y = [N_F - N_G] / [N_G - 1] \qquad (5)$$

In TOS, Y can take any value running from 1 to (almost) infinity [1; ∞] depending on the size of the group, the number of fragments in the increment $N_{F\,\mathrm{inc}}$, in relation to the size of the whole lot, which in turn can be characterized by its totality of fragments, N_F.

For the general case of sampling a lot which is *large* compared to a single, or a few (n_{inc}) increments to be aggregated (composite sampling), it was realized that the numerator is essentially N_F and the denominator essentially N_G. Thus, in this case

$Y = [N_F - N_G] / [N_G - 1]$ can be *well approximated* by $[N_F / N_G]$ (if $N_F >> N_G$ and $N_G >> 1$)

This latter expression opens up for a powerful understanding. Physically [N_F / N_G] is the average number of fragments in a group from the lot in question. This approximation makes it possible in certain situations to deal with all potential groups as represented by such an average group (as concerns size). In this way Y = [N_F / N_G] is a measure of the *size* of the groups in question, i.e. the practical increments, used to sample the lot. This is the key theoretical link to the important role of the size of the increments used in the actual sampling operations. It then matters very much ‘who’ or ‘what’ decides this practical sampling tool size; there is a lot of experience in TOS regarding this critical issue, see e.g., [15]. An introduction to heterogeneity and appropriate sampling modes can also be found in [13], which topic is greatly expanded below.

Below the new formulation is similarly focused on the increment size, but now in relation to the general lot size, particularly in the case where the lot no longer can be considered as large, which lies behind all the the standard assumptions outlined above.

Z = segregation factor is a true phenomenological factor, simply meant to represent the degree of long-range segregation intensity, typically used in a simplistic fashion. Z ranges the interval [0; 1] but will never be exactly equal to either bracket, although it can come arbitrarily close for specific materials, Fig. 1. Gy originally introduced Z in a different context, in which, in addition to represent segregation, it also functioned in a more complicated fashion regarding detailed mathematical reductions, see Gy [3], and Pitard [15]. Suffice here to focus on the physical segregation effect, however.

[Figure 1.]

A key theoretical development was derived in Gy [3], that DH_L is *proportional* to CH_L and the major achievement was that this could be expressed explicitly using the same two factors only:

$$DH_L = \frac{1+Y\,Z}{1+Y} CH_L \qquad (6)$$

Gy [6-9] discusses various issues regarding Y and Z in full detail, issues made more accessible by Pitard [15]. Here it suffices to note, keeping in mind that Z for a given material is always a constant smaller than 1, that to all practical intents and purposes, the smaller the effective size of the sampling increments used, i.e. the smaller Y, will have the effect to <u>reduce</u> DH_L. With a higher number of smaller increments, it will be easier to produce a composite sample with better *coverage* of the full lot heterogeneity. Thus the higher the number of increments, the better to counteract the effects of lot heterogeneity[2].

The other key equation to come out of Gy's detailed formulations relates to the sampling variance stemming from *FSE* and *GSE* specifically. Because of the intimate relationship showed in Eq. 6, it was possible to describe the following also by using only the same two factors Y and Z:

$$\sigma^2_{GSE} = Y \cdot Z \cdot \sigma^2_{FSE} \qquad (7)$$

From this equation it appears that in order to reduce the sampling variance contribution from *GSE*, either Y or Z, or both factors need to be reduced (σ^2_{FSE} is constant for a given material)[3]; Y was treated above, and Z can be reduced by mixing the material to be sampled and/or using a

---

[2] Always subject to complying with FSP, the Fundamental Sampling Principle, which states that increments must be extractable from <u>anywhere</u> in the lot. No exceptions from this principle is acceptable.

[3] While in many situations *Y* and *Z* may be inter-connected, one *can* change the number of increments used for composite sampling for example, and thus reduce the numerical value of *Y* – just as the SUO of mixing will decrease *Z*; reducing *var*(*FSE*) can be achieved by another SUO, crushing.

higher number of increments in composite sampling. See Pitard [15] for the full mathematical details. In passing it is noted that it is always also an option to reduce *var*(*FSE*), although this will require added workload, by reducing the top particle size of the material in question; this will drastically reduce *var*(*FSE*) as it is proportional to the third power of the particle size, see all standard TOS references.

This latter approach means that the material in question is now in a completely different status, it is in fact for all sampling purposes a completely new material, the parameters of which (used in Gy's formula) must all be estimated anew – hence the sometimes significantly increased workload.

2.3.2 The critical role of increment size

When sampling significantly heterogeneous materials, the size of the increment plays a critical role. For clarity, this is here illustrated with a few 1-D lot examples without loss of generality.

Below four principally different types of heterogeneity patterns are presented. 1-*D strings* consist of two different kinds of particles with identical masses; the analyte mass fraction is 0 in matrix particles (black), and 1 in analyte particles (grey). In all cases shown below the average analyte mass fraction is 0.1 (=10 %) for the whole lot, i.e., the whole length of the string. The particles are assumed to have identical masses (1 g) and volumes (1 cm^3). As the consequence of this setup the analyte concentrations expressed either as mass, volume or number fractions are numerically identical.

Fig. 2 first illustrates how simulated *in silico* sampling runs relate to practical sampling.

[Figure 2.]

The four heterogeneous lot types (Fig. 3) are here sampled with increasing increment sizes. The lot size, the number of particles forming the lot, was 10,000 equal size particles in each string. All extracted increments were ‘analyzed’ by simply counting the fraction of the analyte particles, i.e., in these experiments there is no analytical error at all. The resulting variance of these analytical results thus presents the total sampling variance only, s_s^2, which is plotted along the y-axis as function of the increment size (delineated along the x-axis).

Fig. 3 illustrates the resulting relative sampling variance as functions of increasing increment sizes from 1 to 5000 particles. The illustrations in the lower part show the results for model heterogeneity patterns shown in the upper part of the figure. For comparison **B** = shows the theoretical results of a random binomial distribution. Vertical arrows show the effect of segregation as the sampling variance difference in comparison to the ideal case of sampling a *random* binary lot with identical average concentration with the segregated lot. Fig. 4 shows the results as log-log plot in order better to show the relative differences.

[Figure. 3.]

[Figure 4.]

The principal advantage of using a larger increment size is obvious as evidenced by the clear sampling variance reductions revealed (Figs. 3 and 4). This effect varies as a function of the different types of heterogeneities illustrated. The binomial distribution is usually considered as the archetype *random* distribution, which serves well as a reference of an ideal lot. This is always the easiest lot type to sample of all alternatives, but unfortunately, this simplistic model has only very little realism for the overwhelming number of material lots in science, technology and industry.

In Figs. 3 and 4, the vertical double-headed arrows show the effect of *segregation* (and grouping) in comparison to this ideal *random* binary model. There is but little systematic patterns observable in the segregation-induced inflationary effects on the sampling variance, an important indication that ‘heterogeneity’. Even in these carefully constructed simplistic examples segregation effect defies any systematization. The empirical equivalents of these relationships will be more marked for even less well-behaved, more realistic heterogeneous lots materials met in real-world practice.

Lesson: Segregation and grouping, especially if the segregation pattern consist of clusters of analyte containing particles, makes it more and more difficult to obtain a valid and reliable sample rendition of increasingly heterogeneous, irregularly distributed lots. Below it will be emphasized however, that the size of the increment as used here in these experiments are to be understood as total composite sample sizes, i.e., the above is manifested not to be interpreted as blank permission for grab sampling with increasing sample mass. Attempt to counteract the effect of segregation by increasing the size of the individual increments (increasing increment mass) mainly reduces the short-range variance (as shown below), whereas increasing the number of increments reduces the effect of the long-range sampling variance in estimating the lot mean and its variance. Consequently, it is advantageous to collect large(r) samples in all situations using *composite* sampling instead of taking the larger samples as single increments. In practice it is indeed well known that a high(er) number of small(er) increments, n_{inc}, distributed so that they cover the full lot volume as well possible, is by far the most efficient way to improve sampling procedures by better counteracting lot heterogeneity.

2.3.3 Dependence of the between-increments variance on increment size and heterogeneity type

In case the concentration, a_i, or another property of the lot to be estimated, is indeed randomly distributed (ideal case), the sampling variance can be estimated from a pilot study collecting a sufficient number of samples from the lot and calculating their variance, s_s^2. In this case, the variance of the lot mean is inversely proportional to the number of extracted increments, n_{inc}, (classical statistics):

$$s_{a_L}^2 = \frac{s_s^2}{n_{inc}} \qquad (8)$$

Similarly, if the increment size, m_{s1}, that was used originally for estimating the sampling variance s_s^2, is changed to m_{s2} the effect on the new sampling variance can be predicted:

$$s_{s2}^2 = \frac{m_{s1}}{m_2}\, s_s^2 \qquad (9)$$

However, equation (9) is valid if-and-only-if the distribution of the measured property is strictly random within the lot. In the presence of *any* type of segregation, clustering or drift, i.e. changes in concentration with time or location, the situation is always more complicated.

Segregation affects both the estimate of the sampling variance of the lot mean and the between-increment variance estimated by sampling only a fraction of the lot in a principally *unpredictable* way; it is only possible to estimate the effects hereof by an empirical experiment, i.e., a Replication Experiment or a variographic characterization [12]. Fig. 3 showed the sampling variance results of four model lots consisting from 1-dimensional strings of different heterogeneity patterns as function of increasing increment size. Three different segregation types were compared to an ideal random binary distribution. In all cases, the number fraction of analytical particles in the 1-dimensional lots were 10 %. The mass fraction of analyte in analyte containing particles = 1 while that of matrix particles = 0. Assuming that matrix and analyte particles have *identical mass*, the mass fraction of the analyte is identical with the number fraction.

In these simulated sampling runs all strings consisted of 9000 particles with zero concentration and 1000 particles with concentration $a_i = 1$. The average of all four ‘lots’ as mass fraction therefore were $a_L = 0.1$, notably *independent* of the specific heterogeneity pattern depicted.

If the particles are regarded individually, the variance of a string consisting of 9000 zeros and one thousand ones[4] is $s^2 = 0.09$, again *independent* on how the analyte particles are distributed in the string. Thus the relative sampling variance of the individual particles of the lot, i.e. the variance of the *Fundamental Sampling Error*, $s_{FSE}^2 = \frac{s^2}{a_L^2} = 9$. s_{FSE}^2 is the quantitative measure of the *Constitution Heterogeneity*, CH_L. CH_L is an *intrinsic* property of the sampled lot and it does not depend on *how* the particles are geometrically distributed within the lot.

The following general conclusions can be drawn from the simulations shown in Figs. 3 and 4:

Random sampling of individual fragments:

- If the lot to be sampled is *random*, the between-sample variance, as a function of sample size, follows the theoretical values predicted from binomial distribution. In this case, it does not matter if the samples are taken as single increments or by compositing several smaller increments or fragments as long as the sizes of the final samples to be analyzed are equal: The expected sampling variance is the same. In Figs. 3 and 4 the solid line

---

[4] $s^2 = \frac{\sum(a_i - a_L)^2}{N_L}$, $N_L = 9000 + 1000 = 10000$

gives the predictions (between-increment) variance as derived from the binomial distribution. This line shows the effect of *increment size (sample mass)*, i.e. the variance of increments consisting of groups of individual fragments or particles. Grouping and segregation effects are treated more closely in section 5 below.

- In case of segregation, if it is possible to collect a composite sample or samples by randomly picking individual fragments (ideal sampling), this process can be regarded as a virtual mixing of the lot and, consequently, the expected result again follows the random distribution. However, this approach is seldom a feasible procedure in practice, either because the fragments are too small and too many to be picked and analyzed individually, or some part(s) of the lot are not accessible, thus preventing truly random (equiprobabilistic) sampling.

Sampling of groups-of-fragments (increments):

- When increments consist of local *groups*-of-fragments (or particles), segregation affects the between-increment variance in an *unpredictable* way. This between-increment variance is now called *Grouping and Segregation variance*, s_{GSE}^2 which is the quantitative measure of the *Distribution Heterogeneity*, DH_L.

- Figs. 3 and 4 show that s_{GSE}^2 depends both on the heterogeneity pattern and increment size. It can only be estimated experimentally, which may be costly, because then replicate measurements to estimate increment size effects have to be carried out by varying the increment size. This costly and somewhat cumbersome approach is, however, always a feasible way to get to come to grips with the otherwise elusive s_{GSE}^2.

- Experiments carried out using only one single increment size, cannot be used to predict how changing the increment size will affect the s_{GSE}^2, because DH_L depends on the specific heterogeneity pattern of the lot and has to estimated empirically for different increment sizes.

- At extreme segregation, i.e., extreme clustering (heterogeneity pattern 4), DH_L is equal to CH_L.

- Segregation is always the cause of auto-correlation at *some* scale, which has the consequence, as has been shown in numerous cases in the TOS literature (and will also be shown below), that the uncertainty (the sampling variance) of the lot mean depends both on the number of samples analyzed and on the sampling mode. The effect of the sampling mode cannot be estimated without first characterizing the heterogeneity of the lot, which requires however, that one mode of sampling, or other, is used – a vicious circle that cannot be broken. Ramsay [16] suggested the use of a *reference lot* with which to ascertain the effects of alternative sampling modes (and/or the same sampling mode applied with different factors, e.g., different n_{inc}). This approach was criticized severely for lack of clarity of thought and practicality in the light of the above; full details of this critique can be found in Esbensen & Wagner [13].

3 Modified theoretical formulation, *CH* and *DH*

3.1 Constitution Heterogeneity, *CH*

The modified theoretical analysis below relates to the scale of fragments. Let us assume that a lot consists of elementary fragments, each having a mass m_i and a mass fraction a_i of the constituent of interest. The total number of fragments in the lot is N_F, the total mass m_L and the average fragment mass $\overline{m} = m_L/N_F$. Gy defined the heterogeneity contribution h_i of each fragment as the relative deviation from the lot mean a_L[5].

$$h_i = \frac{(a_i - a_L)}{a_L} \cdot \frac{m_i}{\overline{m}} = N_F \frac{(a_i - a_L)}{a_L} \cdot \frac{m_i}{m_L} \tag{10}$$

The true mean concentration of the lot is the *weighted* mean, weighted by the masses of *all* individual fragments of the lot:

$$a_L = \frac{\sum m_i a_i}{\sum m_i} = \sum\left(\frac{m_i}{m_L} \cdot a_i\right) \tag{11}$$

It is important that the lot average is calculated as the weighted mean of all fragments. Especially if the analyte concentration *correlates* with fragment size, the arithmetic mean is biased, Minkkinen [17, 18].

The mean of the heterogeneity contributions is always zero:

$$mean(h_i) = \frac{\sum h_i}{N_F} = 0 \tag{12}$$

---

[5] Analytical results, a_i and a_L are usually given by mass fractions (or mass %). If the results are given as mass-concentrations, the masses in equations 10 and 11 can be replaced as increment or sample volumes.

The variance of h_i is equivalent to the relative variance of the primary increments if these consist of single fragments only, and is then called the *Constitution Heterogeneity* CH_L - it is also called the Fundamental Sampling Error variance σ^2_{FSE} of the average fragment mass $\bar{m}$:

$$CH_L = \frac{1}{N_F}\sum h_i^2 = N_F \sum_{i=1}^{N_F} \frac{(a_i - a_L)^2}{a_L^2} \frac{m_i^2}{m_L^2} \tag{13a}$$

It is of course very seldom possible, or practical, or interesting, to analyze every fragment of a large lot[6].

Instead, it may be possible to collect a sample of n_f fragments and analyze these individually; this is the practice used in so called heterogeneity tests. The validity of this approach is critically dependent upon the assumption that this particular sample is indeed representative of the whole lot. In this case, the fragments have to be collected *individually* using a random, or preferably, a stratified random selection mode in order for this sample to be *representative*.

If the sample is small in comparison to the lot ($n_f << N_f$) then an *estimate*, $\widehat{CH}_L$, of the Constitution Heterogeneity is obtained replacing N_F in Eq. (13a) with n_f -1:

$$\widehat{CH}_L = \frac{1}{n_f - 1}\sum h_i^2 \tag{13b}$$

CH_L is an intrinsic property of any material, *independent* of the spatial distribution of the fragments within the lot, as was shown by the examples above. However, it can be changed, but only if the number and properties of the particles of the lot are changed, e.g., by crushing. Comminution by crushing and grinding changes both the number (N_F) and mass (m_i) of fragments in the lot, as well as the analyte concentration of the fragments, if they consist of a

---

[6] For very special cases, it *may* be of key interest to keep analytical track of all fragments in a lot, e.g. if the fragments are very big relative to the lot size, but there is no generalization possible to the much more often occurring ‘standard’ cases in which sampling in the sense of TOS applies.

mixture of analyte and matrix. Aggregation (agglomeration) of the particles, e.g. crystal-sticking-together or crystal growth in crystallization processes has the opposite effect; N_F decreases while the masses m_i increase, while the lot mass m_L is staying constant.

Both these principal cases, crushing and aggregation, corresponds to a fundamentally new material system, new lots, with completely new heterogeneity characteristics. All features pertaining to the previous system, e.g. estimated heterogeneities, estimated sampling error variances etc. have to be re-estimated. There are no characteristics that are transferable to such a new system.

Multiplying Eq. 13a by the average fragment mass gives an equation for another quantity, which Gy called the *Heterogeneity Invariant*, *HI*, or the Constant Factor of the Constitution Heterogeneity:

$$HI_L = \bar{m} \cdot CH_L = \frac{\bar{m}}{N_F} \sum h_i^2 = \sum_{i=1}^{N_F} \frac{(a_i - a_L)^2}{a_L^2} \frac{m_i^2}{m_L} \qquad (14a)$$

Usually it is not possible, or practical, to analyze all fragments of a lot. However, in that special case, if the fragments/particles are large enough to be analyzed individually, HI_L can be estimated analyzing a sample consisting of n_f elementary fragments:

$$\widehat{HI}_L = \bar{m} \cdot \widehat{CH}_L = \frac{\bar{m}}{n_F - 1} \sum h_i^2 \qquad (14b)$$

HI has the dimension of mass, if the analyte content is given as mass fraction. HI_L represents the variance, the sampling error, of a *virtual sample* having the size of unit mass (expressed e.g. in kg, g, mg) depending in which unit m_i and m_L are given. If the concentration is given as mass/volume, the unit of HI_L is that of volume (dm^3, m^3).

For a particular lot from which HI_L is estimated, the variance of the fundamental sampling error of a composite sample of size $m_s = \sum_{i=1}^{n} m_i$, consisting of *n* randomly picked particles, depending on the sample size of the composite sample, is

$$CH(m_s) = \frac{HI_L}{m_s} = \sigma_{FSE}^2(m_s) \tag{15}$$

Eq. 15 can be used to predict the effect of *changing* the increment or composite sample mass, but again, *only* if the particles are randomly picked from the lot without changing the particle properties (or a composite sample with size m_s is made of particles picked strictly randomly as is the key procedure in heterogeneity tests). Only for this *ideal* sampling process can *FSE* be used to estimate the total sampling variance of the lot mean, $s_{TSE}^2(a_L)$. Eq. 15 gives the variance of the fundamental sampling error of as function of the sample mass picked from a random lot as single increments.

If an estimate, *CH* or *HI* of a lot is available and the lot mean is estimated from n_{comp} composite samples made picking the particles (fragments) individually and each having a mass m_s, the total sampling variance of the lot mean a_L of an ideal sampling process is

$$s_{FSE}^2(a_L) = s_{TSE}^2(a_L) = \frac{HI_L}{m_s \cdot n_{comp}} = \frac{s_{FSE}^2(m_s)}{n_{comp}} \tag{16}$$

Any type of segregation in a stationary lot, or drift of the process mean with time (dynamic, flowing lots), which breaks the randomness of the analyte distribution, will generate *other error components*. For example, process, manufacturing and environmental data sets often show different kinds of periodic or quasi-periodic fluctuations.

The *FSE* calculations given above provides an estimate of the sampling variance of an *ideal* sampling process, which constitute the *theoretical minimum* sampling error of any lot involved. This can only very rarely be achieved in practice however for natural materials and the type of materials of interest in technological and industrial materials processing and manufacturing. In addition, in practice it is of course impossible to study a(ny) lot by picking and analyzing all individual particles or fragments. Instead, resort will always have to be to taking *increments*, consisting of local groups–of-fragments, usually taken with the view of producing composite primary samples.

3.2 Distribution Heterogeneity, DH_L

The much more realistic situation, which dominates in practice over the ideal cases treated above, is that there is distinct non-randomness in the lot. In natural lots there is practically always *segregation* at *some* scale, e.g., due to stratification caused by gravity or centrifugal forces and differences in particle properties, gradual changes in process stream with time (drifting mean); and there are many other agents that can also contribute towards a breakdown of an established mixing uniformity; Pitard [15] discusses these issues in full detail. Analyte particles and fragments may also show a propensity to form more-or-less coherent and well-defined *clusters* of analyte containing particles or grains of different sizes.

Gy [3] defines Distribution Heterogeneity, DH_L, as the between-group (between-increments) variance originating from such grouping and segregation effects.

The theoretical analysis of practical sampling of realistic lot heterogeneities now proceeds at the scale of groups-of-fragments (i.e. the practical increment sampling scale), but derivation of the pertinent heterogeneity characteristics follows the exact same formalism as for the fragment scale.

If the lot is divided into n_g *virtual* groups (*potential* sampling increments, $n_g = m_L/m_n$) having concentration a_n and mass m_n, the distribution heterogeneities of these increments are defined as

$$h_n = \frac{(a_n - a_L)}{a_L} \cdot \frac{m_n}{\overline{m}_n} = n_g \frac{(a_n - a_L)}{a_L} \cdot \frac{m_n}{m_L} \tag{17}$$

Just like the mean of the constitution heterogeneities of fragments, also the mean of the increment heterogeneities is always zero

$$mean(h_n) = \frac{\sum h_n}{n_g} = 0 \tag{18}$$

The *between-increment* variance, called Distribution Heterogeneity, *DH*, is the mean of squared heterogeneities, h_n:

$$DH_L = \sigma^2(h_n) = \frac{1}{n_g} \sum h_n^2 = n_g \sum_{n=1}^{n_g} \frac{(a_n - a_L)^2}{a_L^2} \frac{m_n^2}{m_L^2} \tag{19}$$

If the increment distribution is fully and completely random (or randomized by picking individual increments randomly from the lot), *DH* can be derived from HI_L or CH_L as function of the average increment mass, m_{inc}.

DH_L of *random distribution* is $DH_L = \frac{HI_L}{\overline{m}_{inc}} = \sigma^2_{FSE}(\overline{m}_{inc})$ (20)

If a lot has <u>any</u> type of non-isotropic characteristic; segregation, clustering or location- or time-dependent *changes* in concentration, then DH_L is a linear combination of all short-range and long-range sampling variances.

The short-range variance σ_{SR}^2 has two components: the increment size effect due to the variance of the Fundamental Sampling Error (HI_L/m_s) and the variance due to the segregation *within* of the groups/increments sampled, σ_{Seg}^2. The long-range variance may also have two components: σ_{LR1}^2 and σ_{LR2}^2, in which σ_{LR1}^2 is generated due to a *drift* or other non-periodic change of concentration dependent on locations of the sampling sites in stationary lots (trends), or in sampling time (process sampling, environmental monitoring or similar). σ_{LR2}^2 is the variance due to a possible periodic, or quasi-periodic fluctuation of the concentration within the sampling target or along the process time direction. This type of semi-regular heterogeneity fluctuation is often observed in technological and industrial processing or manufacturing cases.

Thus, for a *non-random distribution* the following applies:

$$DH_L = \sigma^2(h_n) = \sigma_{SR}^2 + \sigma_{LR}^2 = \frac{HI_L}{\overline{m}_n} + \sigma_{Seg}^2 + \sigma_{LR}^2 \qquad (21)$$

If, as above for *CH*

However mixing of the lot before sampling will *reduce* both σ^2_{Seg} and σ^2_{LR}. In Figure 4 the random mixture shows the effect of grouping when the increment size is changed. The difference between the random mixture and the different segregation patterns shows the effect of segregation on the sampling variance (DH_L). As Figure 4 shows, this is a complicated function of sample mass and segregation pattern. Minkkinen & Esbensen, [1] investigated in more depth some of the practical effects of these relationships.

3.3 Binary mixtures: a special case

3.3.1 “Gy’s Formula”

Eqs. 13 and 14 are exact without any assumptions.

For a first approximation to estimate of CH_L and HI_L of materials containing analyte particles *imbedded* in the matrix fragments Gy [2] presented an equation (often called the “Gy equation” or “Gy’s Formula”) with the aim to estimate the relative sampling variance s_s^2 due to the fundamental sampling error (only), as a function of observable characteristics of the lot material. This early achievement quickly became famous in the sampling and other communities, although the originator himself was distinctly unhappy with the many outlandish applications that quickly appeared under his name (*pers. com*), see also [6-9].

The details of the formula need careful attention:

$$s_s^2 = \frac{f\ g\ \beta\ c}{m_s}\ d^3 \tag{23}$$

Here m_s is the mass of the sample increments (either individual fragments for an ideal sample or groups of fragments for a composite sample), and d is the characteristic fragment top-size, or particle top-size, defined as the upper cut-off size that that in sieve analysis retains 5 % of the material.

The shape factor f is defined as the volume ratio of the sampled particles having the characteristic dimension d to the volume of a cube having the same side length d. For spheroidal particles f is approximaed well by 0.5 which is therefore often used as the *default value* for this factor. This is based on Gy's extensive empirical crushing experiments Gy [3, pp. 82-86], where several types of solid materials very often produced ~spheroidal fragments with a shape factor close to 0.5. Well-known exceptions from this are flaky materials, e.g., gold nuggets and mica schist with f equal to [0.1 - 0.2].

g is a characteristic of the size distribution. In Gy's many crushing experiments most materials resulted in a wide size distribution (with ratios of upper and lower cut-off sizes, $d_{0.95}/d_{0.05} > 4$). If the characteristic particle size was defined as upper cut-off dimension, the size distribution factor was close to the value $g = 0.25$ allowing this to be used as another *default* value for approximate evaluation of s_{FSE}^2. A more accurate value *can* be estimated from an empirical size distribution analysis, at the cost of the sieving work normally needed in analyzing large samples for size distribution.

Composition factor c can be estimated if the particle properties of a binary mixture are known:

$$c = \frac{\left(1 - {}^{a_L}/_{\alpha}\right)^2}{{}^{a_L}/_{\alpha}} \rho_c + \left(1 - {}^{a_L}/_{\alpha}\right) \rho_m, \qquad (24)$$

where a_L is the mean concentration of the lot, α the concentration of the analyte in the critical particles, ρ_c and ρ_m are densities of the analyte-bearing particles and matrix respectively. In liberated particle mixture of equal particle sizes, c is equal of the sampling variance of 1 g samples

β = liberation factor, is an *empirical correction factor* used if the analyte bearing particles are found *embedded* as inclusions in other matrix fragments. If the analyte particles are fully *liberated* β = 1.

Many attempts have been made to model the liberation factor with simple empirical equations, e.g. $\beta = (d_{lib}/d)^x = d^{3-x}$. The liberation size d_{lib} is defined as the characteristic size of the embedded analyte particles. Such a simple model is applicable only for a material for which x is estimated empirically and within a relatively narrow fragment size range. See Minnitt [19] for an in-depth introduction to the many intricacies involved in dealing with β, which covers the extensive literature corpus involved well.

Box: Segregation vs. increment size (mass)

The strings shown in Figs. 3 and 4 can also be interpreted as *solid rods* - to represent ores, rocks drill cores etc. Here sampling variances depend both on the sample mass (sampling:= ‘cutting rods into pieces’, or sub-lengths) and the specific heterogeneity pattern. For increments cut from the solid rods, or increments of the same mass taken from the particle strings, the sampling variances are *equal* as long as their heterogeneity patterns are *similar*. From the simulations in Figs. 3 and 4 a fundamental insight regarding the liberation factor (β) appears:

In all four lots all analyte particles are liberated, but only the results for heterogeneity pattern 1 follow Gy’s equation 23 for which $\beta = s_s^2 \cdot m_s/(f\ g\ c\ d^3)$. The three other heterogeneity patterns give liberation size estimates which are larger than 1.00 cm used in the simulations, and different when estimated at different increment size. Assuming full liberation and randomness (β = 1), the characteristic particle size can be determined by reorganizing Eq. 23: $d^3 = s_s^2 \cdot m_s/(f\ g\ c\ \beta)$. When d is estimated from the ideal random set results, correct liberation size estimates are obtained *independent* on increment size (m_s) (line 1 in Figs. 3,4). For the three remaining heterogeneity patterns, the particle size estimates *vary* with increment size. In these cases, the empirically estimated particle sizes are identical to the particle sizes of equivalent binary mixtures (equal analyte and matrix particle sizes) that would give equal sampling variance with the same increment mass. Consequently, using the actual liberation size, the liberation factor β, estimated from the empirical sampling variances, is different from 1. Thus β estimated from the segregated lots represents the *combined* liberation and segregation effects. This feature was a prime motivating driver for the present theoretical developments.

In case of segregation, it is impossible to estimate how changing increment size will affect the sampling variance (in order to do that, extensive (expensive) experiments have to be carried out using several increment sizes). Independent of segregation pattern, when the increment size approaches the liberation size, the sampling variance approaches the theoretical Gy equation curve. While in simulations to it is possible to ‘analyze’ every fragment and increment as was done above, in practical sampling it is only possible to analyze a limited number of increments. Segregation is discussed below in more detail. In simulations it is possible to ‘analyze’ every fragment and increment, in practical sampling it possible to analyze only a limited number of increments.

END OF BOX

.

Gy’s equation has received much more focus in the sampling community and literature than Gy himself intended. It has been vastly misused, but also unduly criticized, forgetting the purpose and the assumptions made in deriving it (the authors humbly confess to having been guilty of the same sin, earlier in their careers – this is not an easy matter to master). This equation was originally derived specifically to allow an approximate estimation to an order-of-magnitude only, of the sampling variance *of binary mixtures*. The liberation factor was introduced to allow the sampling error estimation of unliberated materials; however, it was not understood at the time that the empirically estimated liberation factor is in fact also a measure of segregation, and thus not as simple as originally envisaged. Pierre Gy's own journey is well rendered in Gy [6-9], and in Minnitt [19].

Thus the *assumptions* made in deriving Eq. 23 need being outlined clearly: 1) particles containing the analyte and the matrix particles are assumed to be *similar*, i.e., they have *similar shapes* (or at least similar masses) and *similar size distributions* expressed as mass fractions, 2) matrix particles do not contain the analyte. When Eq. 23 is used to estimate the relative variance of a liberated *mixture* (β = 1), Eq. 23 gives the relative variance of a binary mixture consisting of particles having similar size and shape where the matrix particles do not contain any analyte. This solution is identical with models derived from the binomial distribution.

There is practically no limit to the number of materials in the literature, which (very often *wrongly*) have been forced to 'fit' into these straightjacket assumptions, materials which are de facto very different from binary mixtures. It is best not to give individual references; it is only interesting how to move on towards a more reflected use of Gy's formula.

If *DH* or *HI** have been estimated experimentally, the particle size *d* of a liberated binary mixture giving the *same variance* as the experimental mixture, can be estimated by reorganizing Eq. 17.

$$d = \left(\frac{DH \cdot m_S}{f\ g\ c}\right)^{\frac{1}{3}} = \left(\frac{HI^*}{f\ g\ c}\right)^{\frac{1}{3}} \qquad (25)$$

Given the special assumptions of $f = g = 1$, the characteristic particle size of an *equivalent* binary mixture consisting of cubic particles is obtained. For gold, typical values are: *f* = 0.15; *g* = 0.25. These values are used in the example worked out below.

Figure 5 shows the difference of the *HI* from Eq. 23 in comparison to exact value from Eq. 14a. The results show that when matrix particles are *smaller* than the analyte particles, the difference is small and Gy's equation gives a fair approximation of the *FSE*, but the difference increases rapidly, when matrix particles become *larger* than the analyte containing particles.

[Figure 5.]

Thus before Gy's formula *may* have quantitative relevance as per its original objective (an *order-of-magnitude* estimate *only*), there are usually far more damage done by not focusing sufficiently, if at all, on the prime sampling directive: elimination of all ISE, followed by maximal reduction of the effects stemming from $\boldsymbol{\sigma^2_{GSE}}$. The quantitative effects from these sampling errors very often surpass **those stemming from** $\boldsymbol{\sigma^2_{FSE}}$ alone by orders-of-magnitude themselves**.**

4 Practical sampling of heterogeneous lots

As mentioned, normally it is not possible to analyze all, or a large enough number of individual particles or fragments of any real (large) lots. Only in simulations, or if a small lot is *constructed* in order to demonstrate the sampling process in lab scale, is this possible.

However, CH_L and DH_L can sometimes be estimated experimentally by sampling the lot. For example, if the lot particles or fragments are large enough, so that they can be analyzed *individually*, a reasonably good estimate of CH_L and HI_L *can* be obtained; it is then only a matter of the effort one is willing to put in. Examples of such materials would be pelletized or granulated products, large grain aggregates, certain food commodities (e.g. nuts). Even small individual particles can be analyzed by micro analytical techniques, like electron microscopy with X-ray analysis and in many recent cases using chemical image analytical techniques (CHEMSCAN). Lyman [20] gave a very illuminative illustration of this powerful approach used for smaller and smaller lags. Bedard et al. [21] discussed the critical case of significantly 'nuggety' reference materials used in X-ray analysis with illuminating examples; the reference material masses in typical use were shown to be significantly to small in several cases.

DH_L can also be estimated analyzing increments consisting of groups fragment (or particles) sampled from the same lot. Some form of stratified composite sampling mode should be used in order to have a good representation of the full lot volume.

4.1 Short-range and long-range variance vs. increment size

1. If increments consisting of liberated particles are analyzed individually, the manifestation of both long-range and the short range-variance depends on increment size.
2. The sampling variance of a *random lot*, in which the sampling increments consist of single particles, is the between-particle variance, σ_s^2 and can be estimated from the lot heterogeneity: $\sigma_s^2 = \frac{HI_L}{\bar{m}_{inc}}$.
3. If analyte particles within a *heterogeneous lot* are distributed as forming more or less irregular patterns or trends (collectively called *segregation* in the treatment below), and if increments consist of more than one particle, the sampling variance now depends on the increment size chosen, because segregation affects both the within-increment variance as

well as the between-increments variance. In this case, the sampling variance, i.e., the between-increments variance now has three components:

$$\sigma_S^2 = \frac{HI_L}{\bar{m}_{inc}} + \sigma_{Seg}^2 + \sigma_{LR}^2.$$

$\frac{HI_L}{\bar{m}_{inc}}$ is the part of the sampling variance arising from the constitution heterogeneity and is the ultimate smallest sampling variance that can only be achieved, if the lot can be completely *randomized* before sampling, $[\frac{HI_L}{\bar{m}_{ninc}} + \sigma_{Seg}^2]$ is the within-increment variance or *short-range variance*, σ_{SR}^2 [σ_{Seg}^2 is the part of segregation variance due segregation *within* the increments] and σ_{LR}^2 is the contribution of the within-lot segregation to the complete sampling variance.

The short-range variance (σ_{SR}^2) cannot be predicted from the constitution heterogeneity of the lot. The within-increment variance can be calculated from the analytical results stemming from analyzing all fragments/particles within an increment; which is possible in simulations (as is done here) but seldom in practice (unless fragments are big, or few, relative to the whole lot).

The situation is slightly more complicated because of potential *in-situ heterogeneity*, i.e., irregularly distributed analyte micro-particles embedded within fragments. If these are liberated, "grinding to completion", this will add to the observable short-range variance, but if grinding stops before these micro-particles are liberated, their effect will be hidden, i.e. will be *included* in the short-range variance σ_{SR}^2.With experimental CH_L and DH_L (or *HI*) values, it is now possible to test (approximately) if the sum of the short-range segregation or within fragments variance and the long-range variance is significant *in comparison* to the CH_L:

$$F = \frac{\bar{m}_n\, DH_L}{CH_L} \approx \frac{CH_L + \bar{m}_n(\sigma_{Seg}^2 + \sigma_{LR}^2)}{CH_L} \qquad (26)$$

If this F-test is significant, then the subtraction $DH_L - CHI_L/\bar{m}_n$ gives the estimate of the sum $\sigma_{Seg}^2 + \sigma_{LR}^2$ for sample size $\bar{m}_n$. This F-test is only approximate however, because it assumes that numerator and denominator arise from independent normally distributed distributions. A better, *modified* F-test based on variographic experiments has been presented by Minkkinen [22].

Unlike the constitution heterogeneity CH_L, the distribution heterogeneity DH_L is always affected if the lot is manipulated/agitated e.g. by shaking, transportation, or by deliberate mixing. Also, if the increment size is changed, DH_L changes too, but unlike the case of random distribution, this effect and its effect on the variance of the estimate of the lot mean, estimated analyzing just a part the lot cannot be predicted based on simple experiments with a narrow increment size range. Further, the effects of *GSE* are manifestly *transient*, i.e. varying without any tractable means.

Lastly, in the case of the distributional heterogeneity, the variance of the mean of several increments *depends on the sampling mode*. Only if the n_{inc} increments forming the lot are collected *randomly*, following TOS' rules of correct (unbiased) sampling can the sampling error (variance of the lot mean) be estimated as $s^2(h_n)/n_{inc}$.

However, the effect of sampling mode can be estimated by analyzing the results of a variographic experiment [3, 15, 18].

4.2 Practical Example: estimation of the segregation variance

The data analyzed in the example below are part from a study published by Pitard [23]. In this experiment, a large composite sample (340 kg) of gold bearing ore with an average content of 30.6 g/kg was made by picking 50 increments from a single type of mineralization, subject to the Fundamental Sampling Principle (FSP). The entire composite sample was crushed to -2 cm and divided into four sub-samples by using fractional shoveling. In the Pitard's experiment the data of which is analyzed here, one of these sub-samples was further divided into four sub-splits. These were each crushed or ground to four different nominal particle sizes and subsequently divided again with a rotating sample divider into 16 sub-samples, which were pulverized to -0.106 mm. From each of these 16 pulverized samples, a 60 g composite sample was composed by 24 randomly picked increments.

[Figure 6.]

Finally, these 60 g samples were divided into two 30 g halves, which were all analyzed by fire assay for gold. In total number of fire assays was 128.

It pays to be attentive of the design for this experiment, Fig. 6, as it illustrates several of the key stipulations for representative sampling.

From each of the final duplicate analysis results, the mean and variance were calculated. These variance estimates have just one degree of freedom, but *pooling* them by calculating their mean a variance estimate s_1^2 with 16 degrees of freedom is obtained. Another variance estimate s_2^2 with 15 degrees of freedom is obtained calculating the variance of the mean values of the duplicates.

Of these s_1^2 is the estimate of the total analytical error variance including the splitting of the 60 g sample into duplicate 30 g samples submitted for the fire assay.

$$s_1^2 = s_{TAE}^2 \tag{E1}$$

s_2^2 has three components: Long-range or between-increments variance, short-range or within-increments variance and, because duplicate analyses were made on each sample, the total analytical variance divided by 2.

$$s_2^2 = s_{LR}^2 + s_{SR}^2 + \frac{s_{TAE}^2}{2} = s_{Seg}^2 + \frac{s_{TAE}^2}{2} \tag{E2}$$

The significance of the segregation can now be tested:

$$F = \frac{2 \cdot s_2^2}{s_1^2} = \frac{2 \cdot s_{Seg}^2}{s_{TAE}^2} \tag{E3}$$

If *F* is not significant, we can accept that $s_{Seg}^2 \approx 0$. If it is significant, we can estimate the segregation variance:

$$s_{Seg}^2 = s_2^2 - \frac{s_{TAE}^2}{2} \tag{E4}$$

From these calculations, we get the total variance of single analysis of the 60 g samples

$$s_{Tot}^2 = s_{Seg}^2 + s_{TAE}^2 \tag{E5}$$

TABLES 2a – 2d

The laboratory that carried out the analyses made public its own estimate of the relative standard deviation of the fire assay as 2 % ($s_{TAE}^2 = 0.02^2 = 0.0004$). In this example, all 60 g samples were analyzed twice (using the 30 g duplicates completely in fire assay) giving the estimate of the total sampling variance:

$$s_{samp}^2 = DH_L = \ s_{Tot}^2 - \frac{s_{TAE}^2}{2} = s_{Seg}^2 \qquad \text{(E6)}$$

The analytical results and the estimation of s_1^2 and s_2^2 for all four nominal particle size groups are presented in Tables 2a to 2d and summary of all calculation results in Table 3.

TABLE 3.

When the material was ground to nominal particle size *d* = 0.15 mm, the variance estimate s_2^2 does not differ significantly from s_1^2, which indicates that at this particle size class the Au nuggets in practice are liberated to such a degree that after this last crushing, the fraction of particles still embedded in fragments with a size below 0.15 mm, did not result in a statistically different variance estimate. In short, the crushing cascade has reached the practical liberation size.

If the nuggets were randomly distributed in the matrix, the expected number of Au nuggets/increment would be the same in case the increment sizes are equal as in this experiment (60 g), i.e. the expected number of nuggets randomly distributed in the matrix depends only on the increment size. As this is not the case, this is a clear indication of *segregation* so that some fragments contain clusters of nuggets or some exceptionally large nuggets. This is quite natural, since the majority of gold nuggets are found on the crystal boundaries of the matrix minerals.

The nugget particle size in a binary liberated mixtures, where gold nuggets and matrix particles have similar masses and where the sampling variance is equal with this mineral mixture used in the experiments, where estimated by using Gy’s equation (Eq. 23):

$$d = \left(\frac{s_{rSeg}^2 \cdot m_S}{f\, g\, c}\right)^{\frac{1}{3}} \qquad \text{(E7)}$$

The following values were used: $f = 0.15$; $g = 0.20$; $c = 5.3 \cdot 10^5$ g/cm^3 and $m_s = 60$ g for the three coarsest size classes. As in the finest size class, the segregation was not significant, $\boldsymbol{s_{r1}^2}$ gives the contribution of the fundamental sampling error this size class. The sample mass 30 g were used in estimating the nugget size, 0.148 mm.

The variance of a random particle mixture is inversely proportional to the expected number of particles of interest in the sample. Therefore, if the sample size is kept constant, the expected variance should not change whether the analyte particles are embedded in the fragments or liberated Minkkinen [18]. On the other hand, if the variance increases with increasing fragment size it is a clear indication of some form of segregation of the analyte in the lot. Another indication of segregation is that predicted particle size by using Eq. E7 is smaller than the fragment size and changes with the fragment size.

There are important lessons to be learned from the present re-analysis of the data from the Pitard [22] experimental data:

1. Segregation, whether within-fragment (short-range) or long-range variation of concentration or within the original sampling target, will affect the between increments variance. Consequently, it is impossible to estimate without extensive

experiments how changing the fragment and increment size will affect the segregation manifestations.

2. Full characterization of the heterogeneity pattern in 2-D and 2-D targets is expensive. However, when fragment size is close to liberation size, the predicted particle using Gy's equation is close to the liberation size of the analyte-containing particle.
3. The so-called heterogeneity tests (HT) are popular in the mineral processing arena. However very great care has to be taken in interpreting and utilizing HT results in planning subsequent sampling plans. The reason is that standard HT's do not give information as to the *type* of heterogeneity present, only the magnitude, nor of possible autocorrelation. If experimental HT results are used to estimate a required minimum sample mass, or maximum particles size, in order to be within the required sampling uncertainty decided upon, the results are not reliable unless random or stratified random sampling modes are used with particle and increment sizes *identical* to those that were used in the heterogeneity test itself.

5 Grouping and segregation factors, *y* and z – *new derivation*

When increments are taken from a lot, every removed increment in principle changes the average concentration of the remaining part of the lot, the more the larger the removed increments are in comparison to the whole lot. This effect is of course not significantly influential as long as the extracted increment mass is *much smaller* than the lot mass, i.e. under normal primary sampling in science technology and industry.

However, if size of the increments removed are significant in comparison to the lot size (small lots on the laboratory bench, e.g., in connection with sample preparation or *bona fide* riffle, where the resulting sub sample is half of the lot splitting; in the pharmaceutical industry example exist, where the entire product lot of a highly specialized, extremely expensive compound was of the order of seven grams) this effect has to be taken into account when the short-range grouping effect is estimated, Sommer [24].

If the analyte distribution in the lot is *random* and increments of n_f fragments are extracted (single fragment, group of (or grab) of several fragments or a composite sample is made of n_f randomly picked individual fragments or groups of fragments), the increment mass and number effects on the lot mean can be predicted. The lot mean (a_L) is calculated as the mean of the extracted n_{inc} increments and its variance is

$$\sigma_{a_L}^2 = \frac{\sigma_s^2}{n_{inc}} \frac{N_L -}{N_L - 1} = y\, \sigma_s^2 \tag{27}$$

σ_s^2 is the sampling variance of the increment, which can be estimated empirically analyzing the individual increments or, if the necessary data is available, from heterogeneity $\sigma_s^2 = \sigma_{FSE}^2 = HI_L/m_{inc}$.

From this the definition of our proposed *modified* grouping factor, y, is

$$y = \frac{1}{n_{inc}} \frac{N_L - n_{inc}}{N_L - 1} \tag{28a}$$

The average fragment mass is m_f, N_L is the total number of fragments in the lot and n_{inc} the number of fragments in the single increment or in the composite sample: $N_L = \frac{m_L}{m_f}$, total sample mass $m_s = n_f\, m_f$. Using these relationships the following equality can be derived, in which these numbers are replaced by the equivalent masses. Now y is expressed as a function of the masses of the lot, increment and fragment, respectively.

$$y = \frac{1}{n_{inc}}\left(\frac{m_f}{m_f}\right)\frac{N_L - n_{inc}}{N_L - 1} = \frac{1}{n_{inc}}\frac{m_f \cdot N_L - m_f \cdot n_{inc}}{m_f \cdot N_L - m_f} = \frac{1}{n_{inc}}\frac{m_L - m_{inc}}{m_L - m_f} \qquad (28b)$$

If $n_{inc} = 1$, and m_{inc}, is a multiple of m_f Eq. 27 with y calculated from Eq. 28b gives the dependence of the sampling variance (variance between analyzed increments) as function of increment mass.

The following approximations can be made from Eqs. 28 under the given conditions:

1) If the whole lot is taken as the sample, $m_L = n_{inc} \cdot m_{inc} = m_s$, $m_L - m_s = 0$ and $y = 0$, and, consequently, also the sampling error is zero.

2) If $m_L \gg m_f$ then $m_L - m_f \approx m_L$ and $y \approx \frac{1}{n_{inc}}\frac{m_L - m_{inc}}{m_L}$. (28c)

In Table 4 this approximation is compared to the exact formula. The table shows that this is an applicable approximation in most practical cases.

3) If the lot is *large* in comparison to the total sample size (the usual situation in sampling),

$$m_L \gg m_{inc} \text{ and } m_L \gg m_f,\ y \approx \frac{1}{n_{inc}}, \qquad (28d)$$

which leads to an assumption used in standard statistics: The variance of an increment (composite sample) is the variance of the fragments divided by the number of fragments.

Table 4.

The results given in Table 4 show that only a small error is made by using this approximation when the increment size is *less* than, say, 2 % of the lot. This is the case in most practical situations when primary sampling is carried out.

But when laboratory sub-samples are made, especially at the end of the full ‘lot-to-aliquot’ pathway, typical mass reduction is carried out splitting contemporary sub-samples into 2 -10 parts only (lower part of Table 4), of which one, or a few, are selected as aliquots to be analyzed. In these cases, it is critical to consider the grouping effect correction.

If the heterogeneity invariant for a particular material to be sampled can be estimated, the theoretical variance of an increment is

$$s_{inc}^2 = \frac{HI_l}{m_{inc}} \frac{m_L - m_{inc}}{m_L} \qquad (29)$$

Figures 3 and 4 showed the effect of different heterogeneity types on the sampling variance. For a given increment size (m_{inc}), DH_L is the corresponding sampling variance.

Following the formalism laid down in standard TOS (outlined above), by multiplying the grouping factor with the *modified* segregation factor z (still representing the segregation status of the material), the following equality is valid:

$$DH_L = z \cdot y \cdot s_{FSE}^2 \qquad (30)$$

From this follows an *alternative formulation* of the segregation factor as

$$z = \frac{DH_L}{y \cdot \sigma_{FSE}^2} = \frac{DH_L}{y \cdot CH_L} \qquad (31)$$

If we compare Eq. 30 with Gy's original equation (Eq. 23), we see that by using this new definition, the liberation factor β is formally equal to the new segregation factor *z*. The difference to Gy's liberation and segregation factor is that *z* explains the combined effect of liberation and segregation. Now, if a reference variance (s_{FSE}^2) can be estimated, either theoretically from known properties of a particle mixture or experimentally, e.g., grinding or crushing a solid material close to the liberation size of the analyte containing particles, *z* can be estimated as function of the increment or fragment size.

If the increments sampled are larger than single analyte particles, Gy's equation gives a general equation of the distribution heterogeneity that can be written for *random mixtures* as function of the sample mass:

$$DH_L(m_s) = \frac{HI_L}{m_s}\left(\frac{1}{m_s} - \frac{1}{m_L}\right) = f \cdot g \cdot \beta \cdot c \cdot d^3 \left(\frac{1}{m_s} - \frac{1}{m_L}\right) \tag{32}$$

If the particles are fully liberated, and there is no segregation, $\beta = 1$ and

$$DH_L(m_s) = f \cdot g \cdot c \cdot d^3 \left(\frac{1}{m_s} - \frac{1}{m_L}\right) = y \cdot \sigma_{FSE}^2 \tag{33}$$

In case there is any type of segregation, $\beta \neq 1$ and this is taken account by the new segregation factor:

$$DH_L(m_s) = z \cdot f \cdot g \cdot c \cdot d^3 \left(\frac{1}{m_s} - \frac{1}{m_L}\right) = z \cdot y \cdot s_{FSE}^2 \tag{34}$$

In the gold experiment example above m_s = 30 g and m_L = 16·30 g results in the grouping factor value $y = \frac{1}{m_s} - \frac{1}{m_L} = 0.3125\ g^{-1}$. When the ore was ground to the nominal particle size *d* =

0.015 cm that particle size group did not show significant segregation, so $DH_L = s_{r1}^2 = 0.00118$ in that group and $s_{FSE}^2 = s_{r1}^2/y = 0.0377$ (Eq. 27). For the other groups $DH_L = s_{rSeg}^2$ and, from Eq. 26, $z = s_{rSeg}^2/(y \cdot s_{FSE}^2)$. That gives following values for the segregation factor in the other size groups: *d* = 0.135 cm, **z = 2.38**; *d* = 0.43 cm, **z = 7.00**; and *d* = 1.35 cm, **z = 5.81**.

If the distribution of gold nuggets in the matrix rock is random, the predicted sampling variance due to pure grouping effect is $s_{group}^2 = y\, s_{FSE}^2$. If/when there is segregation, the observed sampling variance is *larger* than predicted. To obtain the segregation effect the grouping variance has to be multiplied by segregation factor *z*. **The segregation factor *z* thus gives the ratio of segregation effect to pure grouping effect.** This constitutes a physical framework for interpreting the different ***z***'s in the different size groups in the above gold sampling experiment; there are *different* contrasts of grouping and segregation manifestations in the different size classes. It is very, very difficult to crush irregular, originally strongly heterogeneous materials to a comparable status in decreasing size class ranges, especially when the analyte is made up of malleable gold particles.

6 Estimation of variance of the lot mean

In standard statistics, the variance of the lot mean of *n* samples analyzed is obtained by dividing the measurement variance s_{tot}^2 by the number of samples:

$$s_{a_L}^2 = \frac{s_{tot}^2}{n} \qquad (35)$$

This relationship holds *only*, if the distribution of the analyte is truly random.

If the lot material is segregated, the experimental variance estimate, e.g., DH_L, *depends* on the size of the sampled increments. Consequently, Eq. 35 is valid depending on two conditions:

1) The increment size is identical to that used in estimating the measurement variance, and

2) The lot was *randomized* before sampling (or virtually randomized by taking the increments randomly so that each potential increment of the lot had an equal chance to be extracted as a sample).

Another point to remember is that when there is segregation there is also auto-correlation *between* the increments, at least within some distance (*range*). In auto-correlated series the variance of the lot mean, i.e. the sampling error, depends on the sampling mode (random, stratified or systematic). Variography is a powerful approach to estimate the effect of the sampling modes in autocorrelated series, Minkkinen [18, 22]. Esbensen et al. [26, 28] and Minkkinen et al. [27] present a comprehensive illustration of these features for the case sampling for GMO quantitation from port offloading in which all degrees of ship's cargo heterogeneities was experienced. This example illustrates a reference case(s) in which analyte distribution is anything but uniform, the KeLDA GMO study [26-28] is a very realistic, didactic case history narrowly paralleling the present treatise.

If the analyte distribution is random however, the variance of the increments can be explained by the grouping effect alone (there is no segregation), i.e. $DH_L = y \cdot CH_L$ and from this follows that at this extreme, for completely random (i.e., non-segregated) lots, z = 1. At the opposite end of the spectrum of heterogeneity, corresponding to extreme segregation (Fig. 3, heterogeneity pattern 4), the distribution heterogeneity is equal to the constitution heterogeneity and therefore here $z = 1/y$.

With respect to TOS' original *Y* and *Z* definitions, only the numerical running intervals for these factors are changed, while their phenomenological meanings are the same:

	Original TOS			Modified TOS (present)
Y	(grouping factor)	[1; ∞]	y	$[1; \frac{1}{n_{inc}} \frac{N_L - n_{inc}}{N_L - 1} \approx \frac{1}{n_{inc}}, if N_L \gg n_{inc}]$
Z	(segregation factor)	[0; 1]	z	$[1; (1/Y \approx n_{inc}, if N_L \gg n_{inc})]$

Between these extreme end values, the new segregation factor *z* depends on the actual distribution, as shown in Figs. 3 &, 4.

The grouping factor *y* can be estimated for materials consisting of specific particles, if the mixture composition is known using the theoretical sampling error models, e.g. Gy's equation, (Eq. 23). That is the case, e.g. in laboratory when mixtures are made for calibrating analytical instruments and in some other cases where small sampling targets are investigated.

7 Theory vs. practice

While *CH*, *DH* and *HI* are theoretically necessary concepts in TOS, their use in practice is limited. Before expensive sampling campaigns on large stationary lots are designed, pilot studies (e.g., replication experiments) on smaller scales in order to optimize the final sampling plan are strongly recommended. From the results of a well-designed pilot study, *DH* can be estimated reliably. If the sampling plan is made in order to estimate the mean concentration of the lot, the grouping and segregation factors, *y* and *z*, and the estimated variance of the lot mean are valid only if random (or stratified random) sampling is used, and the increment size used is the same as in the pilot study.

However, in sampling small lots like in preparing analytical samples from the primary samples *CH*, *DH* and *HI* are useful concepts.

If segregation, due clustering, linear drift and/or cyclic variations within the lot is a significant component of the sampling variance, that is a clear indication of auto-correlation in the lot, at least within some time interval or distance range. In auto-correlated series, the measurement uncertainty, the variance of the mean, *depends* on both the sampling mode, the increment size and number of increments analyzed. For such, significantly heterogeneous lots, the only resort is to empirical total sampling error estimation (*TSE*), by either Replication Experiments or variographic characterization [12, 13]. This can then be guided by the theoretical derivations above.

As always, it is strongly advisable first to reduce, or eliminate fully, all Incorrect Sampling Error (ISE) effects so that the critical sampling variance estimates can become valid and reliable, lest the effects of the *inconstant sampling bias* will dominate unduly, and in most cases make all the estimations outlined above irrelevant [3,4,6-8,10,13,15].

In a recent “refutation” of Esbensen & Wagner [13], Ramsay [29] managed to read most of the Theory of Sampling in a manifest negative and ill-informed fashion, disregarding all of the above heterogeneity vs. sampling mode inter-dependencies - with a fatal result. The present paper can also be seen as a fundamental underpinning of a countermand to [29], which will be published separately elsewhere. In this context also the broad mathematical modelling reflections offered by Francois-Bongarcon [30] plays a central role.

8 Conclusions

1. There is complete command over all sampling errors only for an ideal lot with a *random* distribution of the analyte, subjected to ideal sampling, i.e., extraction of one fragment at the time, independently, with free access across the entire lot volume. In this ideal case the only sampling error is σ_{FSE}^2 which is a bona fide estimate also of DH_L. This is the only case for which the original "Gy's Formula" was derived. The binomial and Poisson distributions offer some insight in this case, but for two components only [analyte, matrix]. Crucially there is no possibility for generalization to any other, more realistic lot type(s). This point is often overlooked.
2. There has been an ill-reflected, often unrecognized, extensive abuse of Gy's Formula during the entire history of applied TOS, it being all too liberally applied to almost any aggregate material conceivable (many material classes of widely different compositions with significant (to large, or extreme) fragment size distribution heterogeneity. This abuse regimen is for the most part characterized by the lack of fundamental TOS understanding and competence; the most recent misconstruction of TOS is in [29]. The present paper is a strong warning against this practice, as is Francois-Bongarcon [30].
3. In any type of a realistic lot, non-randomness (segregation, grouping, linear drift and/or cyclic variations) will produce significant effects stemming from the complementary *GSE*. Possible combinations of *GSE* effects, stemming from a specific sampling mode interacting with a specific lot heterogeneity, are so many [1]) so many that *GSE* cannot be encapsulated in an easy mathematical dressing (like for *FSE*). In dynamic lots and lot which are manipulated the *GSE* effects are transient, which further complicates attempts to generalize the estimation of their magnitudes.

4. For significantly heterogeneously lots (all the world's realistic sampling targets) only empirical *TSE* estimations are possible, e.g., a Replication Experiment (for stationary lots) or a variographic characterization (for process sampling).
5. For sampling in which the intended sample mass is *close* to the total lot mass (2-10%, or more) the presently derived *grouping factor correction term* (exact: $\frac{m_L - m_{inc}}{m_L - m_f}$ and its approximation: $\frac{m_L - m_{inc}}{m_L}$) is critically necessary and mandated.
6. The modified *y* and *z* factors still function in a manner similar to the phenomenological factors *Y* and *Z* in standard TOS, but the *z* factor now represents the combined, more realistic effect of liberation and segregation combined. We submit this gives an easier path into the core intricacies of certain details of TOS needed when more than standard primary sampling is on the agenda, i.e. when the sampling increment size is, or has to be, a substantial fraction of the total lot mass. This is always the case at the terminal end of the 'lot-to-aliquot' pathway and which also characterizes some so-called 'sampling cells' solutions offered in Process Analytical Technologies (PAT). Another advantage is, as the worked out examples showed, that using *y* and z the combined effect of the liberation and segregation can be quantitatively estimated from sampling experiments as function or the increment size. Estimates of HI_L or CH_L are needed and these also can be estimated experimentally, or theoretically from the known composition of the material to be sampled.

Acknowledgements

P M wants to express his gratitude to Lappeenranta University of Technology for allowing to use the facilities of the university in his research at Professor emeritus status. The authors thank profusely the estate Holm Moelle, Denmark for support on so many levels – all hail to Dorthe and Niels Kjaer.

References

[1] P. 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 (2009) 59–70.

[2] P. M. Gy, Erforderliche Probemenge-Kurvetafeln (Minimum sample mass—Graphs and Curves). Internationales Kongress für Erzaufbereitung. Third International Mineral Processing Congress, Goslar, Germany, May 1955. Erzmetall, 8, (1955) B 199–220.

[3] P. M. Gy, Sampling of Heterogeneous and Dynamic Material Systems, Elsevier, Amsterdam, 1992.

[4] P. M. Gy, Sampling for Analytical Purposes, John Wiley & Sons Ltd, Chichester, 1998.

[5] R. C. A Minnit, The cost of sampling errors and bias to the mining industry, TOS Forum, issue 6 (2018) (in press).

[6] P. M. Gy, Sampling of discrete materials – a new introduction to the theory of sampling I. Qualitative approach, in: K. H. Esbensen, P. Minkkinen (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 Practise, Chemom. Intell. Lab. Syst. 74 (2004) 7-24.

[7] P. M. Gy, Sampling of discrete materials – II. Quantitative approach - sampling of zero-dimensional objects, in: K.H. Esbensen, P. Minkkinen (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 Practise, Chemom. Intell. Lab. Syst. 74 (2004) 25-38.

[8] P. M. Gy, Sampling of discrete materials – III. Quantitative approach - sampling of one-dimensional objects, in: K.H. Esbensen, P. Minkkinen (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 Practise, Chemom. Intell. Lab. Syst. 74 (2004) 39-47.

[9] P. M. Gy, Part IV: 50 years of sampling theory – a personal history, in: K.H. Esbensen, P. Minkkinen (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 Practise, Chemom. Intell. Lab. Syst. 74 (2004) 49-60.

[10] F. F. Pitard, and D. Francois-Bongarcon, Demystifying the Fundamental Sampling Error and the Grouping and Segregation Error for practitioners, Proceedings World Conference on Sampling and Blending 5 (WCSB5), Santiago, Chile, 2011, pp. 39-56.

[11] G. Matheron (author), F. F. Pitard, and D. Francois-Bongarcon (translators), Comparison between samples with constant mass and samples with constant fragment population size [and calculations of their sampling variances], in: K. H. Esbensen and C. Wagner C. (Eds.) Proceedings of the 7th World Conference on Sampling and Blending. TOS Forum 5 (2015) 231-238.

[12] DS 3077. Representative sampling - Horizontal Standard. Danish Standards. www.ds.dk, 2013.

[13] K. H. Esbensen, C. Wagner, Theory of Sampling (TOS) versus Measurement Uncertainty (MU) – a call for integration, Trends Anal. Chem. 57 (2014) 93-106.

[14] K. H. Esbensen, Materials Properties: Heterogeneity and Appropriate Sampling Modes, J. AOAC Int. 98 (2015) 269-274. http://dx.doi.org/10.5740/jaoacint.14-234, in: K. H. Esbensen, C. Paoletti, N. Theix, Representative Sampling for Food and Feed Materials: A Critical need for Food/feed Safety. J. AOAC Int. 98, (2015).

[15] F. F. Pitard, Pierre Gy's sampling theory and sampling practice (2nd edn), CRC Press LLC: Boca Raton, USA, 1993.

[16] M. Ramsey, Contaminated land: cost-effective investigation within sampling constraints. In: P. Potts, M. West (Eds.), Portable X-ray Fluoresence Spectrometry: Capabilities for In Situ Analysis. RCS Publ., 2008, pp. 39–555. ISBN 978-0-85404-552-5.

[17] P. Minkkinen, Weighting error – is it significant in Process analysis? in: J. F. Costa and J. Koppe, eds., WCBS3, Third World Conference on Sampling and Blending, 22-25 October, 2007, Porto Alegre, Brazil, Conference Proceedings, Publication Series Fundacão Luis Englert No. 1/2007, (ISBN 987-85-61155-00-1), pp. 59-68.

[18] P. Minkkinen, Dependence of the variance of lot average on the sampling mode and heterogeneity type of the lot, in: S. C. Dominy and K. H. Esbensen (Eds.) Proceedings of the 8th World Conference on Sampling and Blending, 8 – 11 May, 2017, Perth, Australia, The Australian Institute of Mining Metallurgy Publication Series 2/2017, pp. 319-330. ISBN 978 925100 56.

[19] R. C. A. Minnit, A Generalized form of Gy's Equation for Gold Ores –Empirical evidence, in S C Dominy and K H Esbensen Proceedings of the 8th World Conference on Sampling and Blending, 8 – 11 May, 2017, Perth, Australia, The Australian Institute of Mining Metallurgy Publication Series 2/2017, pp. 331-342. ISBN 978 925100 56.

[20] G. J. Lyman, Variograms: properties and estimation, Proceedings, 6th World Conference on Sampling and Blending, 19-22 November, 2013, Lima, Peru pp. 185-206.

[21] P. Minkkinen, Properties of Process Variograms, Proceedings, 6th World Conference on Sampling and Blending, 19-22 November, 2013, Lima, Peru pp. 241-248.

[22] P. Bedard, K. H. Esbensen, S-J Barnes, Empirical Approach for Estimating Reference Material Heterogeneity and Sample Minimum Test Portion Mass for “Nuggety” Precious Metals (Au, Pd, Ir, Pt, Ru), Anal. Chem. 88 (2016) 3504-3511. **DOI:** 10.1021/acs.analchem.5b03574

[23] F. F. Pitard, Effect of residual variances on the estimation of the variance of the Fundamental Error, Chemom. Intell. Lab. Systems, 74 (2004) 149-164.

[24] K. Sommer, Probenahme von Pulvern und körnigen Massengütern, Springer Verlag, Berlin, 1979, pp. 85-126.

[25] K.H. Esbensen and P. Mortensen, (2010) Process Sampling (Theory of Sampling, TOS) – the Missing Link in Process Analytical Technology (PAT), in Bakeev, K. A. (Ed.) Process Analytical Technology. 2.nd Edition, Wiley, 2016, pp. 37-80. ISBN 978-0-470-72207-7

[26] K. H. Esbensen, C. Paoletti, P. Minkkinen, Representative sampling of large kernel lots – I. Theory of Sampling and variographic analysis, Trends Anal. Chem. 32 (2012) 154-165. doi:10.1016/j. trac.2011.09.008.

[27] P. Minkkinen, K.H. Esbensen, C. Paoletti, Representative sampling of large kernel lots – II. Application to soybean sampling for GMO control. Trends Anal. Chem. 32 (2012) 166-178. doi:10.1016/j.trac.2011.12.001.

[28] K. H. Esbensen, C. Paoletti, P. Minkkinen, Representative sampling of large kernel lots – III. General Considerations on sampling heterogeneous foods, Trends Anal. Chem. 32 (2012) 179-184. doi:10.1016/j.trac.2011.12.002.

[29] M. H. Ramsay, Appropriate Sampling for Optimised Measurement (ASOM) rather than the Theory of Sampling (TOS) Approach, to ensure Suitable measurement Quality: A Refutation of Esbensen & Wagner, Geostand. and Geoanal. Res., 40 (2016) 571-581. doi: 10.1111/ggr.12121.

[30] D. Francois-Bongarcon, Mathematical Modelling and TOS, Proceedings 5th World Conference on Sampling and Blending (WCSB5), October 25-28, Santiago, Chile, 2011, pp. 407-416. ISBN 978-956-8504-59-5.

TABLES: Sampling of particulate materials with significant spatial heterogeneity II: Theoretical re-evaluation of correct sampling errors (*FSE*, *GSE*)

Table 1. Governing Principles (GP) and Sampling Unit Operations (SUO) - the necessary and sufficient framework for all representative sampling (Danish-Standards-Foundation [12], Esbensen and Wagner [13]).

Sampling Governing Principles (GP)

1. Fundamental Sampling Principle (FSP)
2. Principle of Sampling Scale Invariance (SSI)
3. Principle of Sampling Correctness (PSC)
4. Sampling Simplicity (primary sampling + mass reduction)
5. Principle of Lot Dimensionality Transformation (LDT)
6. Lot Heterogeneity Characterization (0,1,2,3 D)

Sampling Unit Operations (SUO)

6. Composite Sampling
7. Comminution
8. Mixing/Blending
9. Representative Mass Reduction (sub-sampling)

TABLE 2a. Gold example of two 30 g splits from a 60 g sample ground to nominal particle size d =1.35 cm. Data part from Pitard's experiment [23].

Sample #	**SAMPLE A (ppm)**	**SAMPLE B (ppm)**	**Mean([A*i*;B*i*]) (ppm)**	**var([A*i*;B*i*]) (ppm^2)**
1	29.92	27.86	28.89	2.12
2	28.45	26.06	27.26	2.86
3	33.73	31.49	32.61	2.51
4	29.09	26.37	27.73	3.70
5	30.99	33.0	32.00	2.02
6	27.86	26.73	27.30	0.64
7	29.65	29.57	29.61	0.003
8	33.94	34.51	34.22	0.162
9	32.59	31.5	32.05	0.594
10	28.64	30.37	29.51	1.50
11	32.14	32.26	32.2	0.007
12	32.09	32.85	32.47	0.289
13	37.58	37.78	37.68	0.020
14	27.94	28.24	28.09	0.045
15	30.63	25.06	27.84	15.51
16	30.2	30.65	30.43	0.101
AVERAGE (ppm) and variance s_1^2 (ppm^2)			**30.62**	**2.005**
Variance s_2^2 (ppm^2)			**9.055**	

TABLE 2b. Gold example of two 30 g splits from a 60 g sample ground to nominal particle size d = 0.43 cm. Data part from Pitard's experiment [23].

Sample #	**SAMPLE A (ppm)**	**SAMPLE B (ppm)**	**Mean([A*i*;B*i*]) (ppm)**	**var([A*i*;B*i*]) (ppm^2)**
1	34.26	33.08	33.67	0.696
2	32.05	30.65	31.35	0.980
3	32.28	32.62	32.45	0.058
4	31.99	31.14	31.56	0.361
5	31.15	29.8	30.47	0.911
6	34.98	29.38	32.18	15.68
7	31.19	32.13	31.66	0.441
8	32.01	30.87	31.44	0.650
9	31.05	30.75	30.9	0.045
10	31.23	33.54	32.39	2.668
11	38.64	34.36	36.5	9.159
12	33.11	32.69	32.9	0.088
13	33.08	33.15	33.11	0.0025
14	34.43	32.87	33.65	1.217
15	33.35	31.74	32.55	1.296
16	32.01	32.83	32.42	0.336
AVERAGE (ppm) and variance s_1^2 (ppm^2)			**32.45**	**2.162**
Variance s_2^2 (ppm^2)			**2.143**	

TABLE 2c. Gold example of two 30 g splits from a 60 g sample ground to nominal particle size d = 0.135 cm. Data part from Pitard's experiment [23].

Sample #	**SAMPLE A (ppm)**	**SAMPLE B (ppm)**	**Mean([A*i*;B*i*]) (ppm)**	**var([A*i*;B*i*]) (ppm^2)**
1	32.41	32.39	32.4	0.0001
2	32.19	29.81	31.0	2.832
3	29.78	32.84	31.310	4.682
4	29.99	28.58	29.285	0.994
5	33.13	34.54	33.835	0.994
6	32.09	30.22	31.155	1.749
7	29.88	28.55	29.215	0.884
8	32.78	31.97	32.375	0.328
9	28.98	28.85	28.915	0.0084
10	32.41	30.7	31.555	1.462
11	32.98	31.33	32.155	1.361
12	32.76	30.97	31.865	1.602
13	33.64	31.4	32.52	2.509
14	35.43	39.62	37.525	8.778
15	31.48	32.54	32.01	0.562
16	31.5	31.22	31.36	0.0392
AVERAGE (ppm) and variance s_1^2 (ppm^2)			**31.78**	**1.799**
Variance s_2^2 (ppm^2)			**4.370**	

TABLE 2d. Gold example of two 30 g splits from a 60 g sample ground to nominal particle size d = 0.015 cm. Data part from Pitard's experiment [23].

Sample #	**SAMPLE A (ppm)**	**SAMPLE B (ppm)**	**Mean([A*i*;B*i*]) (ppm)**	**var([A*i*;B*i*]) (ppm^2)**
1	34.2	33.71	33.99	0.120
2	33.99	33.55	33.77	0.097
3	32.76	32.63	32.70	0.0084
4	34.49	33.91	34.2	0.168
5	36.41	33.95	35.18	3.026
6	31.03	33.54	32.29	3.15
7	31.69	34.34	33.02	3.511
8	33.09	32.31	32.7	0.304
9	34.2	33.14	33.67	0.562
10	33.73	33.62	33.67	0.0060
11	32.92	34.26	33.59	0.898
12	34.63	32.7	33.67	1.862
13	33.18	35.01	34.10	1.674
14	32.6	35.35	33.98	3.781
15	31.66	33.67	32.66	2.020
16	33.16	32.75	32.96	0.084
AVERAGE (ppm) and variance s_1^2 (ppm^2)			**33.51**	**1.330**
Variance s_2^2 (ppm^2)			**0.590**	

TABLE 3. Summary of variance estimate calculations. Variance and standard deviation values are given as absolute and relative values. P(F) gives the cumulative probability of F-distribution at the test value; d_{est} is the estimated particle size of an *equivalent* binary mixture of gold nuggets and matrix having the same sampling variance with the gold ore used in the experiments.

Qantity	**Nominal Particle Sizes (mm)**				**Average**
	13.5	**4.3**	**1.35**	**0.15**	
Average (ppm)	30.62	32.45	31.78	33.51	32.09
s_1^2 **(ppm^2)**	2.005	2.162	1.799	1.330	1.824
s_{r1}^2	0.00185	0.00194	0.00159	0.00118	0.00164
s_{r1} **(%)**	4.30	4.41	3.99	3.44	4.05
s_2^2 **(ppm^2)**	8.451	2.000	4.078	0.551	
s_{r2}^2	0.00779	0.00180	0.0036	0.00049	
s_{r2} **(%)**	8.82	4.24	6.00	2.21	
F-test	8.43	1.85	4.53	0.83	
P(F)< p	0.999	0.88	0.998	0.36	
s_{Seg}^2 **(ppm^2)**	7.45	0.919	3.1 8	≈ 0	
s_{rSeg}^2	0.00686	0.000826	0.00281	≈ 0	
s_{rSeg} **(%)**	8.28	2.87	5.30	≈ 0	
$s_{r(tot)}^2$	0.0870	0.00277	0.00440	0.00108	
$s_{r(tot)}$ **(%)**	9.33	5.26	6.62	3.29	
d_{est} **(mm)**	1.67	0.205	0.692	0.148	

Table 4. Dependence of the grouping factor correction term (exact: $\frac{m_L - m_{inc}}{m_L - m_f}$ and approximation: $\frac{m_L - m_{inc}}{m_L}$) on the relative size of increment taken from the lot. Correction application field (grey).

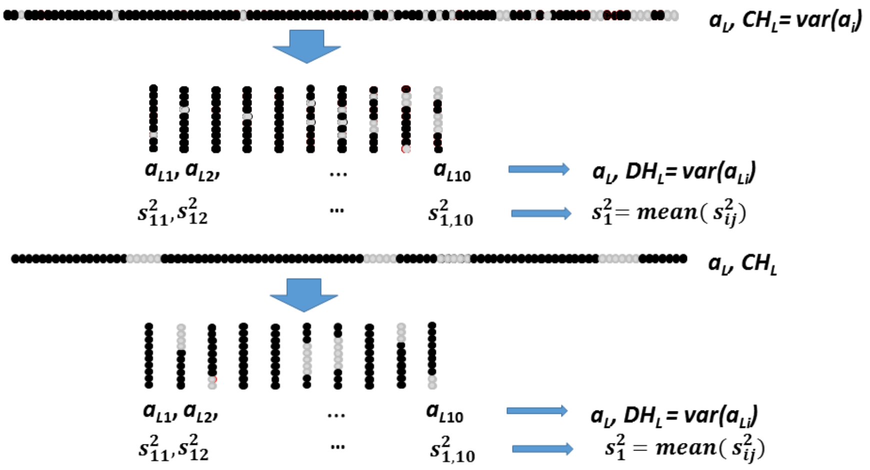
$\boldsymbol{\frac{m_{inc}}{m_L}}$	$\boldsymbol{m_L = 100}$		$\boldsymbol{m_L = 1000}$		$\boldsymbol{m_L = 10000}$	
	$\boldsymbol{\frac{m_L - m_{inc}}{m_L - m_f}}$	$\boldsymbol{\frac{m_L - m_{inc}}{m_L}}$	$\boldsymbol{\frac{m_L - m_{inc}}{m_L - m_f}}$	$\boldsymbol{\frac{m_L - m_{inc}}{m_L}}$	$\boldsymbol{\frac{m_L - m_{inc}}{m_L - m_f}}$	$\boldsymbol{\frac{m_L - m_{inc}}{m_L}}$
0.001			1.0000	0.9990	0.9991	0.9990
0.005			0.9960	0.9950	0.9951	0.9950
0.01	1.000	0.9900	0.9910	0.9900	0.9901	0.9900
0.02	0.9899	0.9800	0.9810	0.9800	0.9801	0.9800
0.05	0.9596	0.9500	0.9510	0.9500	0.9501	0.9500
0.1	0.9091	0.9000	0.9009	0.9000	0.9001	0.9000
0.2	0.8081	0.8000	0.8008	0.8000	0.8001	0.8000
0.3	0.7071	0.7000	0.7007	0.7000	0.7001	0.7000
0.4	0.6061	0.6000	0.6006	0.6000	0.6001	0.6000
0.5	0.5051	0.5000	0.5005	0.5000	0.5001	0.5000

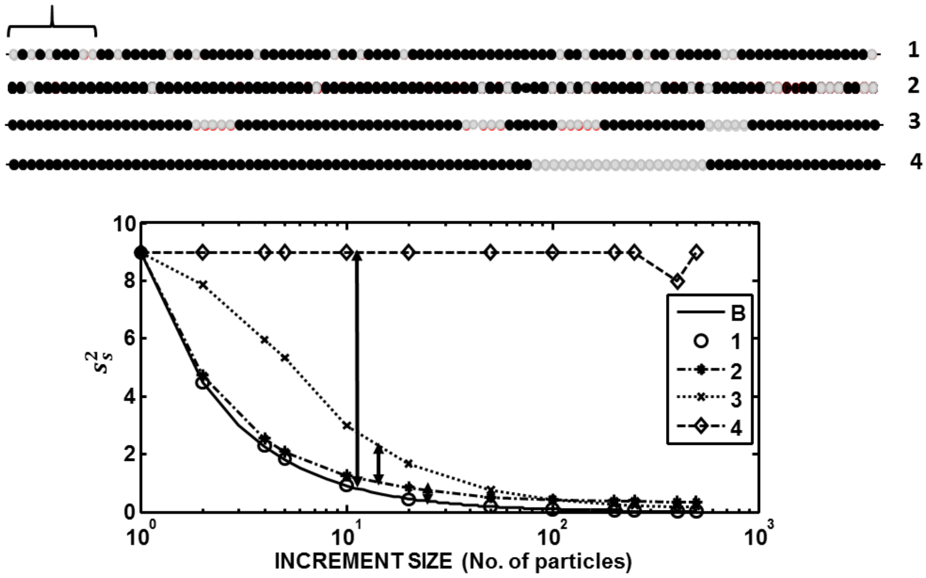
Reducing contributions to total sampling variation from the Grouping and Segregation Error (*GSE*)

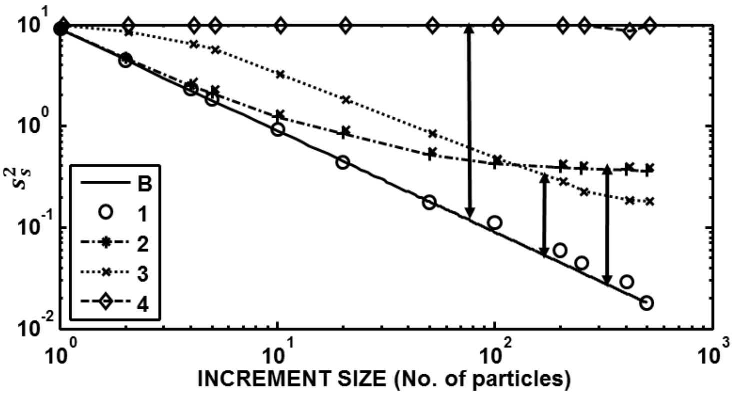
$$s^2(GSE) = Z \cdot Y \cdot s^2(FSE)$$

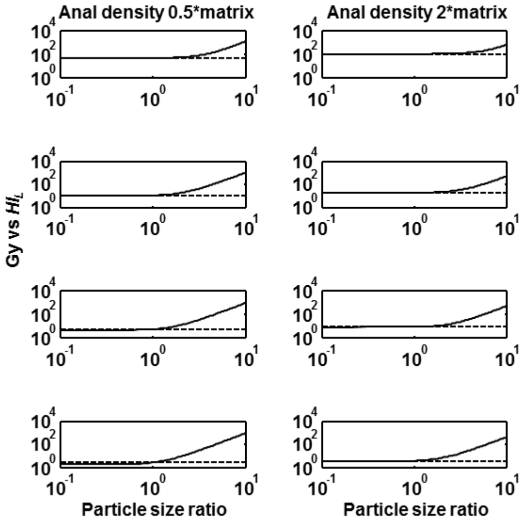
Segregation factor

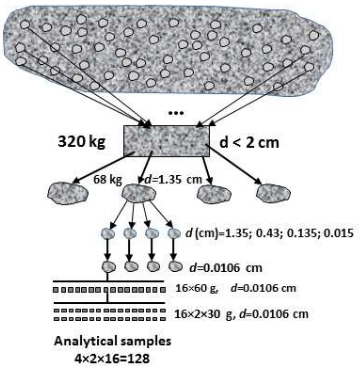
Grouping factor (unaffected
by mixing - reduced only by
selecting smaller increments)


a_L, CH_L= var(a_i)
a_{L1}, a_{L2}, … a_{L10}
a_L, DH_L= var(a_{Li})
s_{11}^2, s_{12}^2 … $s_{1,10}^2$
$s_1^2 = mean(s_{ij}^2)$
a_L, CH_L
a_{L1}, a_{L2}, … a_{L10}
a_L, DH_L = var(a_{Li})
s_{11}^2, s_{12}^2 … $s_{1,10}^2$
$s_1^2 = mean(s_{ij}^2)$


1
2
3
4
10
8
6
4
2
0
s_s^2
B
1
2
3
4
10^0
10^1
10^2
10^3
INCREMENT SIZE (No. of particles)


s_s^2
10^1
10^0
10^{-1}
10^{-2}
10^0
10^1
10^2
10^3
INCREMENT SIZE (No. of particles)
B
1
2
3
4


Anal density 0.5*matrix
Anal density 2*matrix
Gy vs HI_L
10^4
10^2
10^0
10^{-1}
10^0
10^1
Particle size ratio
Particle size ratio


...
320 kg
d < 2 cm
68 kg
d=1.35 cm
d (cm)=1.35; 0.43; 0.135; 0.015
d=0.0106 cm
16×60 g, d=0.0106 cm
16×2×30 g, d=0.0106 cm
Analytical samples
4×2×16=128

Highlights

- New theoretical developments regarding segregation (Theory of Sampling, TOS)
- Didactic simulation illustrations
- Updated summary of practical sampling